ESSAYS ON LAND AND LABOR IN URBANIZING CHINA
Yan Wu 2017

ESSAYS ON LAND AND LABOR IN URBANIZING CHINA

Yan Wu

Propositions

1. Rising housing prices are detrimental to manufacturing development in China. (this thesis)
2. China's regional-competition-growth model contributes to growing land conflicts.
   (this thesis)
3. Although e-commerce was initially hailed as a great opportunity for small businesses, it will hinder their growth because of more intense competition from large investors.
4. Globalization contributes to both the decrease of cross-border income gaps and the increase of domestic disparities.
5. The double-blind peer review policy of academic journals will fail when evaluations by reviewers tend to converge towards 'major revisions'.
6. The challenges that capitalism and democracy were facing in 2016 (and probably will face in the near future) can be dealt with by capitalism and democracy per se.

Propositions belonging to the thesis, entitled
"Essays on land and labor in urbanizing China".
Yan Wu
Wageningen, 18 April 2017.

ESSAYS ON LAND AND LABOR

IN URBANIZING CHINA

Yan Wu

Thesis committee

Promotor

Prof. Dr E.H. Bulte
Professor of Development Economics
Wageningen University & Research

Co-promotors

Dr N.B.M. Heerink
Associate professor, Development Economics Group
Wageningen University & Research

Prof. Dr X. Shi
College of Public Administration
Nanjing Agricultural University, P.R. China

Other members

Prof. Dr H. Folmer, University of Groningen

Dr M.J.J. Handgraaf, Wageningen University & Research

Prof. Dr W.J.M. Heijman, Wageningen University & Research

Prof. Dr Xiaobo Zhang, Peking University, P.R. China

This research was conducted under the auspices of the Wageningen School of Social Sciences (WASS).

ESSAYS ON LAND AND LABOR

IN URBANIZING CHINA

Yan Wu

Thesis
submitted in fulfilment of the requirement for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus,
Prof. Dr A.P.J. Mol,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Tuesday 18 April 2017
at 11 a.m. in the Aula.

Yan Wu
Essays on land and labor in urbanizing China,
155 pages.

PhD thesis, Wageningen University, Wageningen, the Netherlands (2017)
With references, with summary in English

ISBN 978-94-6343-134-7
DOI http://dx.doi.org/10.18174/409700

Contents

CHAPTER 1

Introduction

Urbanization is usually considered an important driving force of economic development. It eliminates the distance between people and firms and thereby the transportation costs, and allows firms to attain the so-called agglomeration economies. In the words of Edward Glaeser (2012), "*the invention of cities makes us Richer, Smarter, Greener, Healthier, and Happier*". Urbanization, however, invites growing challenges as well, for instance, air pollution, crime, loss of agricultural land, lack of affordable housing, soaring social disparity, etc. As stated in the World Cities Report 2016, published by UN Habitat (2016), the model of urbanization today has failed in addressing issues such as 'inequality, climate change, informality, insecurity, and [over-] expansion'.

China has been undergoing dramatic urbanization over the past three decades. In 2011 its urban population exceeded its rural counterpart for the first time in China's history. Productivity and employment gains from urbanization have been major driving forces of China's enormous economic growth. In the view of Joseph Stiglitz, urbanization in China, paralleled with technological innovation in America, is one of the "two keys" to world development in the 21st century (Economist, 2014).

Urbanization in China, however, has also been associated with a range of challenges. For instance, urban and rural development is relatively imbalanced, and public discontent with housing affordability and environment quality in cities has been growing in recent years. Of particular interest, major land and labor flows from rural to urban areas have occurred even though major institutional bottlenecks persist at these two markets. Rural land needed for urban expansion (including real estate investment, industrial development, etc.) has to be acquired by local governments before it can enter the urban market, while *de facto* restrictions on access to public facilities still exist for rural laborers working in urban areas and their families. Land disputes, housing price appreciation, and persistent gender wage gaps among migrants are challenges which may at least partly be related to the institutional bottlenecks in these two markets.

This thesis focuses on some of the major challenges in land and labor markets in urbanizing China. In what follows, I start from a brief overview of the available theory on benefits and problems associated with urbanization. I then document some stylized challenges in land and labor markets associated with urbanization in China over the past few decades. Building on the discussion of these problems, I formulate the research objective, together with the specific research questions that each chapter of the thesis is going to investigate. I also give an overview of the research methods and data used in answering the research questions. A roadmap of the thesis is provided at the end of the chapter.

1.1 Problem statement

1.1.1 Benefits and problems associated with urbanization

Urbanization is closely correlated with rising income levels. More developed countries usually have larger urban populations, and vice versa. To explain the positive relationship between urbanization and income levels, economic theories (e.g., economic geography theories, urban economics) have identified several major mechanisms through which urbanization can improve production efficiency and thus income levels. As noted in Brueckner (2011), scale economies (i.e., 'increasing returns to scale') and agglomeration economies (e.g., inter-firm spillovers) are the two main forces that motivate people and firms to concentrate in particular places. Duranton and Puga (2004) highlight three key channels through which urbanization/cities improve economic performance: sharing, matching, and learning. Sharing refers to the share of indivisible public goods, production facilities, etc.. Matching is mainly based on job searching theories: labor markets become larger and *more* diverse (e.g., in skill levels) as city size increases; this can help reduce the job searching cost (for both employers and employees) (see also O'Sullivan 2012). The learning effect of cities mainly refers to knowledge spillovers in cities (see also Glaeser 1999; Glaeser and Maré 2001).

The benefits of urbanization do not come naturally/without cost. To attain the full advantages of urbanization, cities need to address many market and policy issues/failures. Provision of transport infrastructure is often subject to market failure, and its under-provision can cause, for instance, traffic congestion and environment pollution (see, e.g., Becker and Morrison, 1999). Urban sprawl can reduce the benefits of scale and agglomeration economies (see, e.g., Glaeser and Kahn 2004). Government regulations in housing markets and inter-jurisdictional competition in residents and businesses can distort urban development as well. According to Hsieh and Moretti (2015), local government restrictions on housing supplies in (highly productive) US cities hindered local expansion of working forces; the resulting efficiency loss was estimated at 9.5% of US GDP. In addition, provision of affordable houses and public services in cities, education investments in potential new urban residents, labor migration cost, etc., can all limit people's access to cities and thus the sharing of urbanization benefits (Henderson, 2009).

1.1.2 The challenges of urbanization in China

In the 1980s the urbanization rate in China was only 20%. Urban population, reached 690 million in 2011, accounting for 51% of the population—according to the National Bureau of Statistics (NBS). Urban area in 2010 was over 2.3 times its size in 1990. Urbanization in China accelerated since the early 2000s when real estate construction started to boom. Over the period 2003 - 2014, China built up 100 billion square feet of residential real estate (Glaeser et al., 2016). It was estimated that China used more cement between 2011 and 2013 than the U.S. used in the entire 20th Century (see Swanson, 2015).

Urbanization has produced many benefits to Chinese society. Rural laborers migrating to urban areas usually receive higher incomes as compared with rural jobs. Investments in cities provide better infrastructure and larger market size for firms and services. Urban land and labor markets, however, are facing important institutional bottlenecks. In the land (and housing) market, governments have a strong monopoly power; local governments determine the amount of rural land to be converted to urban land as well as the price of land that can be used for urban construction (Tan *et al*. 2011). In the labor market, it is still very difficult for rural migrants to acquire urban *Hukou* (i.e. urban registration), which provides formal accesses to urban public serveries such as public schools and health insurance (see Meng, 2012). In this section I briefly discuss the challenges in these two markets and the spurred academic debates on these issues.

1.1.2.1 Housing price appreciation, rising living and production cost and deindustrialization

A commercial housing market in China was only set up in the late 1990s. It was suppressed during the planned economy era (see chapter 5 for background). The increase in housing prices has been extremely fast since then. As shown in Figure 1 (blue line), the residential property price in 2000 was approximately 2,000 yuan per square meter for the country as a whole. It reached about 3,600 yuan in 2007. After a temporary adjustment in 2008 (likely due to the global financial crisis), the housing price reverted to fast growth between 2009 and 2014 and amounted to almost 6,000 yuan in 2014. In terms of growth rate, Figure 1.1 indicates the average growth rate was about 8% for the whole period, while there was considerable volatility. The NBS housing price data do not take housing quality into account. Wu et al. (2012) produce an inflation and housing quality corrected housing price index. Their results suggest that the real, constant quality, housing price grew more than a factor of two between 2000 (1st quarter) and 2010 (1st quarter).

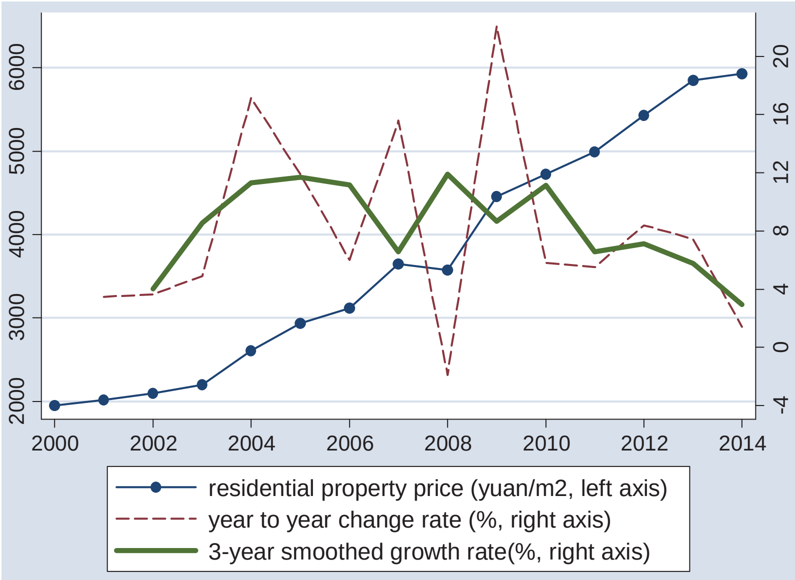


Figure 1.1 Residential property price and its change rates during 2000-2014

Source: Derived from NBS, China Statistical Yearbook (2001-2015)

A direct implication of rising housing prices for households is the increase in living cost (for home purchases or rentals), which raises the pressure on wages for hiring workers. In addition, rising housing prices will raise production cost when firms rent properties or lease new land for establishing production facilities. Consequently, housing price growth can spill over into the whole economy — in a similar manner as a natural resource price boom can spill over. In this regard, a central concern of the relevant literature is the extent to which increases in resource prices cause a decline in manufacturing investment and/or deindustrialization (Corden and Neary, 1982). Housing price appreciation, however, may also have positive effects on firm investments through direct inter-industry linkages — rising demand for construction — or through financial channels — borrowing against increasing housing values (Chaney et al., 2012). Given the important role of industrial employment in China (now and in the near future) and the significance of its real estate price boom, it is crucial to understand how and to what extent real estate prices has affected manufacturing investment in recent years.

1.1.2.2 Real estate investment led growth model and resource misallocation

The real estate sector has become one important driving force of China's economic growth. Between 2003 and 2014, China built 5.5 million apartments on average each year; the newly constructed floor space is equivalent to 74 square feet per person in China (Glaeser et al., 2016). While China has for some time been the world's predominant export factory, it has recently become the world's largest construction site as well.

The real estate boom has raised concerns about resource allocation efficiency. In the first- tier cities such as Beijing, Shanghai, Shenzhen and the second-tier provincial capital cities, business and industries need to cope with high property cost when expanding. As found in Rong et al. (2016), housing price appreciation in these cities stimulates firms to enter the real estate industry; the involvement in real estate activities may crowd out firm's innovation investment. The third- and fourth- tier cities faced slower housing price growth but a fast increase in their housing stock (Glaeser et al., 2016). It is a big challenge for these cities to attract people or business to enter there. Otherwise, the vacant apartments in these cities are a direct waste of resources.

1.1.2.3 Land acquisition and land conflicts in China

Another issue that attracts much attention is the large number of land disputes in China. Although reliable data sources and accurate quantitative numbers on land conflicts are limited, it is well accepted across media, the general public, social scientists and governments that land triggered disputes have become a major cause of social instability in contemporary China. There have been upward trends in land related petitions, massive 'mass incidents' (i.e. protests involving more than 100 participants) and illegal cases (e.g., Lu et al., 2012; Meligrana et al., 2011). It is also observed that land disputes are more likely to occur at the urban edge and are followed by land acquisition; land grab (usually by local authorities) with insufficient and unfair compensation is a typical cause of tension over land.

The prevalence of land disputes is a relatively unexplored field in the literature. The lack of systematic data on such disputes is probably one important reason (see the discussion in section 4.2 of Chapter 4). Also, any data on disputes may still be biased: for instance, information collection is more difficult when governments are less transparent; and social unrest is more closely monitored and heavily suppressed where there tend to be more 'true' disputes. Nonetheless, a small literature has investigated the rationale behind land disputes in China and concluded that land institutions in China do not provide sufficient protection of farm land as farmers do not personally own in rural areas. Moreover, local governments are increasingly aggressive in obtaining land revenues through land requisition from farmers and leasing out the land at much higher prices. According to Chen and Kung (2016), total land leasing revenue for county level governments in 2008 was 1,221 times its size in 1998. As shown in Figure 1.2, cities such as Chengdu, Hangzhou and Hefei even raised more revenues from land leasing than they obtained from formal budgetary items, like value added tax and income tax, over the period 2003-2011. In addition to the fiscal revenue incentive of local governments, studies also emphasize the promotion incentive of local official under China's political system as an underlying reason for rural land grabs: to achieve high economic growth rates and get promoted, local officials often offer land at low prices to attract foreign direct investment (Tao et al. 2010).

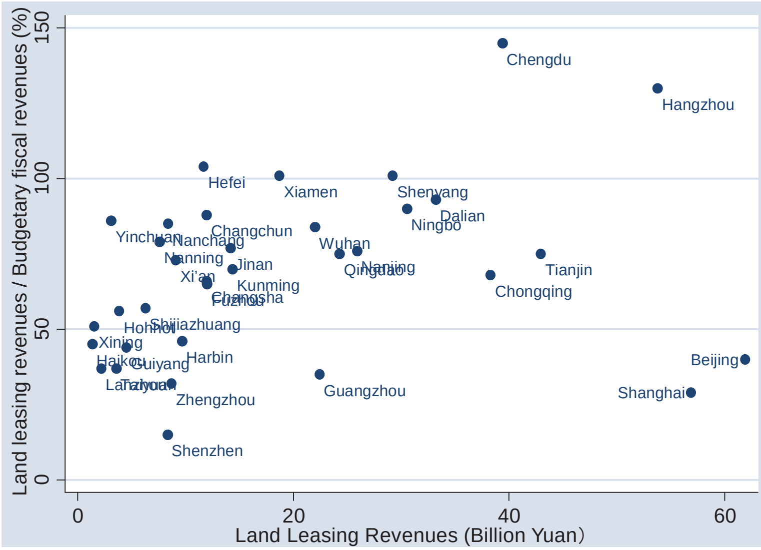


Figure 1.2 Land leasing revenue and its relative size compared with budgetary fiscal revenue (annual averages, 2003-2011)

Source: Derived from NBS, China Land and Resources Almanac (2004-2012) and NBS, China City Statistical Yearbook (2004-2012)

1.1.2.4 Rural-urban segmentation, gender discrimination and urbanization

Along with the rapid urbanization over the past decades, the growth of off-farm income has been a major source of income growth for rural households. Income gaps between rural-urban migrants and urban residents (i.e., residents with urban *Hukou*) have not been closed. For example, the (raw) wage differential between urban workers and migrants amounted to 48% in 2002 and 58% in 2007, respectively (Zhu, 2015). Moreover, only a small share of the pronounced income gap can be explained by labor attributes such as differences in human capital, suggesting discrimination against migrants (see Meng, 2012).

An under-explored aspect of rural-urban migration is gender wage discrimination. It is well-known in China that fewer rural women migrate to cities compared to men. Obviously, the prospect of female migrants (and children) in cities is crucial for rural families settling down in cities. However, rigorous analyses of gender pay differentials among rural-urban migrants in China are scarce. The few available empirical studies on gender wage gaps among rural-urban migrants use data sets that focus on migrants residing in urban communities, and largely exclude migrants living at their workplaces even though that group is known to be sizable in the Chinese context. For these reasons, more attention to the gender perspective is needed for examining rural-urban migration and urbanization in China.

1.2 Objective and research questions

The overarching objective of this thesis is to improve the understanding of several major challenges associated with the urbanization process in China, thereby focusing on two dimensions: *1) the development of the industrial economy; 2) problems facing specific socio-economic groups.* In doing so, I hypothesize and test various explanations—*why the industrial economy may be hurt by the rising housing prices associated with urbanization; why some social groups gain more/less from rural-to-urban movements of land and labor.* With the objective in mind and the stylized facts sketched in Sections 1.2.1 – 1.2.4 above, I limit my attention to the following four specific questions — each as a stand-alone empirical essay:

1. How do local housing price appreciations impact (manufacturing) firms' investment rates (chapter 2)?
2. How do real estate booms shape firm sorting and thus intra-industry resource allocation (chapter 3)?
3. How do foreign direct investment and fiscal decentralization contribute to land conflicts? (chapter 4)
4. How large are the gender wage gaps among rural-urban migrants and what are their sources (chapter 5)?

1.3 Research methodology and data

The research questions listed above are answered by applying state-of-the-art econometric techniques, like panel data models, GMM procedures, instrumental variable regressions and difference-in-differences method, and to (combinations of) data sets that have recently become available.

In Chapter 2, I combine a large data set for manufacturing firms with city-level housing price data. Data for (manufacturing) firms are from the Annual Survey of

Industrial Firms — collected by a professional team of National Bureau of Statistics (NBS) of China. Despite some challenges with the data set, previous studies prove that its quality is satisfying (see Brandt et al., 2012 and Brandt et al., 2014). Average housing price data in cities are drawn from NBS, China City Statistical Yearbooks for various years. The combination yields a large panel data set for firms across most major cities in China for the years 2003-2007. I employ the data set to explore the relationship between housing price and firm investment rate. In addition, I use the geographical characteristics and land leasing data of each city to capture supply constraints of land. They are used as instrumental variables for housing prices.

Data used in Chapter 3 are also based on manufacturing firm data collected by NBS but are aggregated at the 3-digit industry level (according to China Industry Classification). They refer to the period 2000-2007. Using a difference-in-differences method, I explore how the linkages with real estate sectors impact industrial-level total factor productivity (TFP) dispersions before and after the real estate boom in 2003. Data for the linkage measure are drawn from China's Input-Output table for the year 2007.

In Chapter 4 I use a provincial-level data set for the period 1999 – 2010. Land areas involved in illegal land use cases in each province are obtained from NBS, China Land and Resources Almanac. They are used as indicators of the intensity of land conflicts, i.e., the outcome variable of interest. The cross-product of FDI and fiscal decentralization is included as an explanatory variable in the fixed effects regression model to test the main hypothesis that the interaction between FDI and fiscal decentralization contributes positively to land conflicts.

Chapter 5 explores a newly available rural-urban migrant data. The data are drawn from the Rural-Urban Migration in China (RUMiC) project. It covers both migrants living at their workplace and those living in urban communities. The former group is hardly or not covered in previous studies. I use the data to examine the magnitudes and sources of gender wage gaps among migrants. The study involves estimations of Mincer-type wage model and the decomposition of gender wage differences. The decomposition analyses are conducted with two methods — the well-known Blinder-Oaxaca decomposition method and the newly developed Gelbach decomposition method.

1.4 Structure of the thesis

The thesis consists of six chapters. They are organized as follows.

Chapter 1 offers an introductory discussion of the overarching objective and specific questions of the whole research. Chapters 2—5 are four (self-contained) empirical papers — each addressing one research question. They can be divided into two parts. Chapters 2 and 3 focus on the impacts of the real estate boom on the industrial economy. Chapters 4 and 5 address two socio-economic challenges — land conflicts and gender inequality, respectively. Chapter 2 investigates the effects of housing price appreciation on firm investment. I find a robust negative relationship between local housing prices and investments of manufacturing firms. A detailed examination of the underlying mechanism probes that it is mainly due to the Dutch disease effect of the real estate price boom: housing prices push up wages and other production costs for manufacturing firms and thereby reduce the incentive to invest.

Chapter 3 focuses on the intra-industry resource allocation effects of the real estate boom. The results show that industries intrinsically highly linked with real estate sectors experienced increasing heterogeneity in firm-level TFP, suggesting sorting-in or expansion of less-efficient firms in the industries which are highly linked with real estate sectors.

Chapter 4 examines the surging of land conflicts in recent decades. The results show that the FDI growth rate has a positive and significant impact on the growth rate of illegal land use when there is a high degree of fiscal decentralization. It provides evidence that FDI inflows trigger tensions over land in particular in highly fiscally decentralized provinces.

Chapter 5 focuses on the wage gap between female and male urban workers from rural areas in China. It tests whether the wage differences can be attributed to human capital characteristics, gender discrimination, or other causes. I document a relative small gender wage gap of 16-18%, and find that most of the gender wage gap cannot be attributed to gender differences in observed characteristics. In addition, I find some important differences in determining factors of the gender wage gap between the sub-sample of migrants living at their workplaces and those living at urban communities.

Chapter 6 synthesizes the findings in each chapter, discusses their value-added to related literature, and presents recommendations for policy-making and future research.

CHAPTER 2

Local Effects of Housing Price Appreciation on Industrial Investment: Firm-level Evidence from China

Abstract

The findings on the collateral effect of real estate in China contradict with those documented in the U.S. and Japan: positive effects of housing price appreciation on firm borrowing and investment are hardly found or at most only exist for a sub-group of Chinese firms. To reconcile the mixed evidence, we argue that housing price growth can discourage firm investment by pushing up input cost, the so-called Dutch disease channel. We construct a large data set, covering most of industrial firms and prefectural level cities for the period of 2003-2007, to investigate the relevance of the Dutch disease channel. In order to account for the potential endogeneity of housing price, the variation in local geographical constrains in housing construction and the national movement of interest rate are used as instruments. Our empirical results show that there was a significant and robust negative association between local housing prices and the investment ratio of manufacturing firms in the corresponding cities. In particular, the negative relationship is strongest for firms that are labor intensive, non-state owned, and have few linkages with real estate sectors. We conclude that the Dutch disease channel is one missing puzzle in understanding the impact of real estate price on firm investment.

Keywords: Housing price, Manufacturing Investment, Dutch disease, China

This chapter is about to be submitted to Journal of Regional Science, as Wu, Y., N. Heerink, L. Yu. 'Local Effects of Housing Price Appreciation on Industrial Investment: Firm-level Evidence from China.'

Land and Labor in Urbanizing China

2.1 Introduction

The collateral function of real estate has long been recognized (Barro, 1976, and Bernanke, 1983; among others). The theoretical works of Bernanke and Gertler (1989), Kiyotaki and Moore (1997), and Liu et al (2013) have all shown that positive (negative) real estate value shocks can amplify investment upturns (downturns) due to the collateral function of real estate. The theory is broadly supported by the empirical studies on both the U.S. and Japan. For instance, Gan (2007a, b) shows that a decline in land value in Japan during the early 1990s had a large negative impact on corporate investment, whereas Chaney et al (2012) show that real estate price appreciation in the U.S. over the period of 1993-2007 had a positive effect on the investment of U.S. public corporations.

In spite of the aforementioned literature, it is still possible that the collateral effect of real estate can play little role in determining investment behavior. Conceptually, the premise of a positive association between housing price[1] and investment is that real estate should be of importance in resolving firms' financial needs. It is, however, not always the case. For instance, Wu et al (2015) argue that many firms in China are state owned, which usually have easy access to credits and are thus often financially unconstrained. Moreover, debts can be secured without real estate collateral because banks in China are mainly owned by government and are thus able to use government arms, not necessarily real estate, to ensure debt repayments. In this sense, firm ownership, relationship with banks and (or) governments can play much more important roles than real estate in determining credit accessibility in financial markets such as China. In line with the argument, Wu et al (2015) find no evidence that real estate value appreciation increased firm investment for a sample of listed firms located in 35 major cities in China. Chen et al. (2015), employing data for manufacturing firms in 70 cities, show evidence that the

[1] Throughout this paper, we do not distinguish between housing price and real estate price. Both of them refer to property price consisting of residential and commercial buildings as explained in Section 2.3.

positive effects of real estate price appreciation is significant for private firms but not significant for state owned enterprises (SOE).

In addition to the collateral effect, this study provides a new channel linking real estate prices and firm investment. We argue that real estate price appreciation can cause a decline in investment in a way analogous to the Dutch disease mechanism. Classic Dutch disease studies note that the discovery of natural resources and/or an increase in resource prices can cause a decline in the manufacturing sectors. Following the same logic, real estate boom can drive up production costs, including wages, land rents, and interest rates. In addition, this effect should be more observable in tradable sectors, such as manufacturing industries, whereas service sectors theoretically face less elastic demand and can therefore (partially) absorb the rising production cost by setting higher prices for their goods/services. Consequently, the Dutch disease channel will predict a negative causal association between booming real estate prices and investment, particularly in manufacturing sectors. The study sets to test the Dutch disease mechanism as compared with the collateral effect in the context of China.

Our empirical examination employs a Chinese firm-level data set, consisting of rich annual manufacturing firm information (covered by the industrial surveys of the National Bureau of Statistics of China) and housing prices for the cities where each firm was located during the period of 2003-2007. The quantitative assessments of the relationship between housing prices and the firm investment rate lead to the following main findings. Real estate price has a robust negative effect on firm investment. The negative impact is mainly driven by labor cost growth: the higher labor intensity of an industry, the stronger negative effect of housing price. These results imply the existence of a Dutch disease mechanism.

By contrast, there is overall weak evidence that the collateral effect was at work or was of economic importance. This is evident from the following facts: SOE firms generally invest more than their counterparts as housing prices increase, all else equal. However, theoretically real estate price appreciation would facilitate credit access for constrained firms such as non-SOEs if the collateral effect were at work. Alternatively, we relate financial constrains to firm size (following Rajan and Zingales 2001). Consistent with the prediction of a collateral effect, (financially constrained) small firms increase investment more than large firms, everything else equal. But the magnitude of the collateral effect is much smaller than the Dutch disease effect. Finally, we also find that the adverse impact on manufacturing industries is long-lasting rather than temporary.

Therefore, we conclude that a real estate boom can indeed cause a decline in manufacturing industries. It is mainly due to the increase in labor cost predicted from the Dutch disease mechanism. In contrast, the collateral effect is not pronounced in our study on China, where real estate collateral plays a less important role than, e.g., firm ownership in accessing credits.

This study contributes to the literature in the following aspects. First, the paper outlines two opposite mechanisms linking real estate prices with firm investment. One is a well-accepted collateral channel predicting the appreciation of real estate prices. The other is what we call the Dutch disease channel. The distinction sheds light on why real estate prices can impact firm investment differently in Japan and the U.S. compared with China. Second, our sample covers a very large share of manufacturing firms across most Chinese cities. The choice mitigates concerns about sample selection issues arise from geographic coverage of cities. Last, we tackle the endogeneity issue of housing prices in a conventional way but with novel data. Following the literature, we use both the share of unsuitable land for real estate construction (interacted with national movement in the interest rate) and the supply of urban land (adjusted by local GDP) as the instrumental variables for housing prices. The instrumental strategy isolates factors that might drive firm investment and housing prices together and allow us to make causal inference.

The remainder of the paper proceeds as follows. Section 2.2 provides the theoretical background. Section 2.3 describes the data and their sources. Section 2.4 then discuss the empirical strategy. The main results together with an exploration of robustness and mechanisms are presented in Section 2.5. Section 2.6 concludes the paper.

2.2 Theoretical considerations

Two lines of theoretical studies are related with our work. One strand of literature considers real estate as collateral assets in relaxing financial constraints for firms. As a result, the investment decision of firms is impacted by real estate values. By contrast, the second strand of literature emphasizes that real estate price appreciation can push up (local) production costs. In what follows, we discuss each of them.

2.2.1 Collateral effect

The collateral role of real estate has been discussed in numerous economic studies. The theoretical explanation, as presented in Barro (1976), Bernanke (1983) Bernanke and Gertler (1989), Kiyotaki and Moore (1997), and Liu et al (2013), is that positive real estate value shocks can amplify investment upturn, and negative shocks can amplify downturns. The collateral channel gained vast support until the recent work of Wu et al (2015). Gan (2007a, b) and Chaney et al (2012) find a positive relationship from the two sides: the land value decline in Japan during the early 1990s reduced corporate investment, whereas real estate price appreciation in the US over the period of 1993-2007 had a positive effect on the investment of US public corporations. By contrast, Wu et al (2015) find that real estate value changes have an insignificant effect on listed firms' investment in China. They attribute the role of real estate in accessing credits as secondary. The contrasting evidence casts doubt on the positive role of real estate values in spurring firm investment.

A small group of literature focuses on the investment of small firms, or entrepreneurship, for the same reason: Entrepreneurial activity is expected to increase when credit constraints are eased due to real estate price appreciation. The existing evidence, similar as the above collateral effect, differs between China and developed countries. For example, Black et al (1996) and Schmalz et al (2013) provide evidence that house price growth stimulates the start and expansion of small businesses in the U.K. and France, respectively. Adelino et al (2015) also show that real estate prices promote the growth of small business employment, especially in those that needed little start-up capital in the U.S. between 2002 and 2007. However, Li and Wu (2014) show that the rapid growth of house prices in China during the last decade or so has discouraged entrepreneurship. In their view, although real estate price appreciation can reduce credit constraints, it also causes little savings net from paying housing costs.

2.2.2 Dutch disease effect

The Dutch disease theory originally referred to the fact that the discovery of natural resources and/or increase in resource prices can cause a decline in non-resource-linked industries, particularly manufacturing sectors. The early theoretical contribution of Corden and Neary (1982) establishes two channels for the coexistence of a booming resource sector and lagging manufacturing sector. One is called the resource movement effect. The higher profitability and wage levels in the booming sector will draw investment and labor away from manufacturing industries, i.e., the non-booming sectors. This process will further push up factor prices in general and reduce the profitability of manufacturing industries (noting that the demand elasticity of their output – traded goods – is high and thus prices are inflexible). The other one is called the spending effect on the demand side. With wage growth, more income will be spent on non-tradable goods, i.e., services. Investment and economic structure will thus skew toward non-manufacture sectors[2].

This Dutch disease framework has been extended in understanding the economic structure impacts of broad windfall gains, including aid and remittance.

[2] Van der Ploeg and Anthony (2012) provide a good overview of the Dutch disease literature. In this study, we ignore the institutional channel of Dutch disease emphasized by, e.g., Mehlim et al (2006). It is more relevant for international comparisons, whereas our study uses sub-national data.

For example, Rajan and Subramanian (2011) and Bulte et al (2015) show evidence that (foreign) aid can hinder manufacturing development in both international and domestic contexts, respectively. In our view, the fast expansion of real estate construction and/or appreciation in real estate prices can impact manufacturing industries through similar channels. Undoubtedly, manufacturing industries face increasing competition as more and more money, labor and land is reallocated to real estate construction. Production costs will also be pushed up due to the fast growth in property prices. A few studies provide related evidence of the presence of the Dutch disease effect due to the boom of real estate prices in China. For example, Chen et al (2015) documents that the recent appreciation in housing prices in China only increased the investment of commercial land holding but decreased the non-land investment, e.g., equipment.

2.2.3 Summary of the literature

The collateral effect of real estate has been documented in many studies. However, recent studies using Chinese data raise concerns about the existence of the collateral effect. This might due to real estate in developing countries not being sufficient in resolving financial frictions. In contrast, few studies have paid enough attention to the Dutch disease effect potentially triggered by real estate price appreciation. This is much more relevant in emerging countries such as China where secondary industry accounts for the main sectors of the economy. Because the two effects have opposite effects on investment, the relative size of the two effects will determine the net effect. If the real estate collateral effect is insignificant in China, like Wu et al (2015) found, and the Dutch disease effect is present, there will be a negative association between real estate prices and manufacturing investment. Broadly speaking, if the Dutch disease channel is at work and dominates the real estate collateral channel, real estate price appreciation will discourage investment (Figure 2.1).

The above two channels, however, do not cover all the effects that a real estate boom can have on the economy. For example, there is well-documented positive inter-industry spillover of the real estate sector. The development of real estate may promote infrastructure construction and/or urbanization, which generally introduces an agglomeration economy (e.g., Michaels, 2011). Meanwhile, there are also vast literature emphasize the role of housing cost on labor migration and skill sorting (e.g., Saks 2008; Broxterman and Yezer 2015). It is challenging to test each of these mechanisms individually due to data limitations. We first estimate the overall impact of real estate prices on local firm investment and test the possible channels as best we can.

Real estate price	**Micro Channels**			**Firm investment**
	Collateral effect	Credit access	(+)	
	Dutch disease effect	Factor price	(-)	
	Industry-Region Channels			
	Inter-industrial spillover	Industrial linkage	(+)	
	Agglomeration effect	Scale economy	(+)	

Figure 2.1: Conceptual Framework

2.3 Data

First, the firm-level information is obtained from the Annual Survey of Industrial Firms, conducted by the National Bureau of Statistics of China (ASIF)[3]. The ASIF data set provides detailed firm information, including industry affiliation, location, and all operation and performance items from the accounting statements, such as output, intermediate materials, employment, book value and net value of fixed assets. The survey covers all state-owned enterprises and non-state-owned enterprises with annual sales of five million RMB (or above) throughout China. We are thus able to construct detailed firm-level variables, e.g., investment, our main interest. Meanwhile, firms are continuously traced (as long as they exist and are above the scale of ASIF survey, of course). Firm heterogeneities can be better

[3] The data set has been increasingly used since the contribution of Brandt et al (2012). An introduction of working with ASIF data, including tracing firms over time, available variables in the data, and measurement issues, can be found in Brandt et al (2014).

accounted for than in cross-section data. From a theoretical perspective, the use of manufacturing firm data speaks to our theoretical consideration on the Dutch disease effect. In doing so, we are also able to extend both the empirical and theoretical scopes of the prior literature[4].

Second, the Average Selling Price of commercial building (ASPCB) is used as the proxy of the city-level real estate price. The series are calculated as total sales divided by the total amount of housing square footage. The data are available in the China Statistical Yearbook for Regional Economy. Criticism of the data includes the argument that it makes no attempt to control for quality differences across markets or drift over time (Wu et al, 2014). The most recent efforts to construct property quality adjusted housing price series have been made by Wu et al (2014). The comparison of average selling prices (used in our study) with a quality-controlled hedonic price index (available for 35 major cities) finds that the two series are highly correlated (Zheng and Kahn 2013)[5]. It thus increases our confidence in using the ASPCD data. Given the advantage of geographic coverage of both ASIF and ASPCB data sets (all the prefecture-level cities in China), selection biases could be significantly reduced compared with using data for major cities, for instance.

However, our data also have limitations. The ASIF have no real estate holding information. Thus, we cannot account for differences in the quantity of properties and instead must treat it as exogenous. In this sense, we ignore the heterogeneity of real estate holding at the firm level. Additionally, we choose the period 2003-2007 in the study. The ASIF data used by most of studies is for the period 1998-2007, but the reliable housing price data begins in 2003, following the enactment of land market

[4] The full ASIF data cover all industrial firms, including mining, manufacturing, and public utilities. We exclude firms in the sectors of mining and the production and supply of electricity gas and water, whose investment decision may reflect special factors. Therefore, only manufacturing firms are considered.

[5] An alternative data set, Price indices in 70 Large and Medium-Sized Cities, tries to control the impact of housing quality. However, the data set has been criticized for its lack of variation over time and for being severely downward biased (Deng et al, 2012).

acts.[6-7]

Table 2.1 reports the sample construction process and the resulting sample size. As evident from the column titles, the procedure involves the first step of data cleaning of the raw ASIF data and the second step of matching with city-level data. The raw ASIF data (excluding mining and public utilities sectors as mentioned above) for 2003 have 181,186 observations, indicating there were 181,186 manufacturing firms. The number reached 256,999 in the following year. The fast growth is due to 2004 being a census year; thus, the small firms excluded from annual surveys are present in 2004. For the same reason, the number of firms in 2005, i.e., 249,030, was a little bit smaller than in the previous year. The sample size reached 279,282 and 313,046 in 2006 and 2007, respectively.

As commonly done in prior studies, we clean the data following the procedure suggested by Jefferson et al (2008), among others. This involves discarding the observations with the following criteria: (1) key variables such as total industrial output, industrial added value, fixed assets and employees have negative values; (2) the ratio of the value added to sales is smaller than 0 or larger than 1; and (3) the number of employees is smaller than 8, as most of the improbable values are associated with smaller firms that usually have less reliable accounting systems. The procedure reduces the raw data by 4-6 per cent, leaving us with between 169,924 and 300,829 observations depending on the years.

Last, we match the resulting ASIF data with the city-level housing price data. To do this, we need to know whether the location identifier in ASIF data represents the production activities where they happen. For example, all the sub-establishments/plants operate at the same location where the firms register. Our examination of the ASIF data shows that the share of firms producing in more than

---

[6] The ASIF data can be dated back to 1992, the first available survey year. Most of studies work with the sample 1998-2007 because the firm identifiers became consistent from 1998 onwards and the data after 2007 contain missing or unreliable information (Brandt et al, 2014, Feenstra et al., 2013).

[7] After 2002, the Ministry of Land and Resources required local governments to sell land via public auction, market based land and housing price started to be documented reliably.

one city account for less than 5% of all observations in the NBS data (see also Brandt et al, 2012). We thus choose to drop these firms when merging firm-level data with city-level housing price data[8].

After cleaning and matching, we have observations between 140, 000 and 248,000 for different years. However, this sample includes firms that have only one observation during the period of 2003-2007. To better control for firm fixed effects, the analysis will focus on sub-samples with at least 3-5 consecutive year observations during 2003-2007. The last three columns in Table 2.1 show the corresponding number of firms in these sub-samples. We come back to this issue in Section 2.4.4.

Table 2.1: Number of firms in the ASIF data

Year	(1) Raw data	(2) Cleaning	(3) Matching			
			All firms	Age composition of firms		
				>= 3	>= 4	>= 5
2003	181,186	169,924	140,077	70,708	61,803	48,909
2004	256,999	242,382	181,007	87,585	79,023	64,640
2005	249,030	237,218	195,921	158,189	130,174	64,639
2006	279,282	266,971	224,083	171,242	131,467	64,639
2007	313,046	300,829	248,322	160,674	125,198	64,640

Source: ASIF data (2003-2007)
Notes: To further clean the data, we adopt the data-cleaning procedure suggested by Jefferson et al (2008).

Table 2.2 shows the evolution of housing prices according to ASPCB data set. The mean value of nominal housing price was 1,532 Chinese Yuan, approximately 185 USD, per square meter in 2003. The number reached 2,679 Chinese Yuan (or 352 USD) in 2007. The average growth rate was 10%, with the fastest growth in 2005 (14%). The standard deviations of housing price generally grew faster than the means, resulting in increasing coefficient of variations. This implies that there were vast inter-city differences of housing price growth rates. We can thus conclude that: 1) housing prices overall grew quickly in the sampled cities; 2) the inequality of housing prices also increased significantly during 2003-2007.

[8] Therefore, all the firms have only one plant/establishment in our sample. We thus do not distinguish firm from plants or establishments throughout the paper.

Table 2.2: Housing prices and their rate of change: average across city in 2003-2007

Year	City	Average selling price of commercial building (Nominal)								ΔlnHP(Real)
		mean	sd	cv	p10	p25	p50	p75	p90	mean
2003	217	1532	784	0.51	878	1042	1303	1792	2429	-
2004	217	1703	919	0.54	934	1181	1455	2056	2676	7.2%
2005	217	2027	1115	0.55	1118	1344	1647	2354	3683	14.0%
2006	217	2345	1646	0.70	1241	1489	1861	2614	4221	9.4%
2007	217	2679	1724	0.64	1417	1681	2101	3053	4903	9.6%
03-07(Δ)	217	75%	120%	25%	61%	61%	61%	70%	102%	10.1%

Note: The table presents the average selling price in 217 Chinese cities. The last column reports the rate of change of the CPI deflated housing price index (ΔlnHP), simple average across cities.

2.4 Empirical strategy

2.4.1 Model

To provide an estimate of the impacts of housing price on firms' investment, we employ an econometric model close to Chaney et al (2012) and Wu et al (2015). The specification of the regression equation is as follows:

$$y_{i,j,c,t} = \alpha + \beta * \mathbf{ln}(\boldsymbol{HP_{c,t}}) + \boldsymbol{\gamma X}_{i,j,c,t} + \boldsymbol{\tau Init_{i,j,c}} * \boldsymbol{ln}(\boldsymbol{HP_{c,t}}) + \boldsymbol{\delta IND}_{j,t} + \mu_i + \sigma_t + \varepsilon_{i,j,c,t} \quad , \qquad (1)$$

where i stands for the firm, j for the sector, c for the city and t for the year. $y_{i,j,c,t}$ measures firm investments, defined as the ratio of firm investment to the capital stock at the beginning of the period. $\boldsymbol{HP_{c,t}}$ is the housing price in the city where firm i is located. $\boldsymbol{X}_{i,j,c,t}$ is a vector of firm-level control variables that includes typical determinants distinguished in the investment literature. $\boldsymbol{IND}_{j,t}$ is a vector of industry-level control variables, including the 3-digit industry-level agglomeration index and median values of ROE and CASH for the industry to which firm i belongs. Following Chaney *et al* (2012) and Wu *et al* (2015), we further include the interaction terms between core firm-level variables (i.e., ROA and CASH) and housing price, indicated by $\boldsymbol{Init_{i,j,c}} * \boldsymbol{ln}(\boldsymbol{HP_{c,t}})$, to control for the potential differential responses to housing prices due to pre-boom firm characteristics. Last, the variables μ_i and σ_t capture firm- and year-fixed effects, α and β are unknown coefficients, γ and **δ** are row vectors of unknown coefficients, and $\varepsilon_{i,j,c,t}$ is an error term.[9]

To examine the long-run effect, we use a first differenced regression of the above model based on data at the beginning (2003) and the end of our study period (2007):

$$\Delta^{03-07} y_{i,j,c} = \vartheta * \Delta^{03-07}\ln HP_c + \boldsymbol{\rho}\Delta^{03-07}\boldsymbol{X}_{i,j,c} + \boldsymbol{\pi}\Delta^{03-07}\boldsymbol{IND}_j + \epsilon_{i,j,c} \quad , \qquad (2)$$

A further goal of the empirical study is to examine the possible mechanisms that explain the relationship between real estate prices and firm investment. According to our theoretical considerations, we estimate models of the following form:

$$M_{i,j,c,t} = \alpha + \beta * \ln \boldsymbol{HP}_{\boldsymbol{c},\boldsymbol{t}} + \boldsymbol{\gamma X}_{i,j,c,t} + \boldsymbol{\delta IND}_{j,t} + \boldsymbol{\tau Init} * \boldsymbol{lnHP} + \mu_i + \sigma_t + \varepsilon_{i,j,c,t} \quad , \qquad (3)$$

where $M_{i,j,c,t}$is one of our potential channels described below. The intermediating variables we investigate in this section consist of the debt-to-asset ratio (Debt)[10], finance-cost-to-income ratio (Finance), and labour-cost-to-value-added ratio (Labour). These models are thus very similar to our main linear regression Equation (2), except that the left-hand-side variable is different.

Table 2.3 depicts the definitions of the variables we used and summary statistics. The dependent variable is the investment-to-asset ratio, defined as the net change in fixed assets normalized by capital stock at the beginning of each year. The investment and capital stock data are recovered from book values of fixed assets (reported in ASIF data set) by employing the method proposed by Song and Wu (2012)[11]. Note that the investment rate in the Chinese manufacturing firm sample during the study period is heavily right-skewed with a median (9.05%) that is only about 55 per cent of the mean (16.48%).

The key explanatory variable of interest, housing prices, is measured by the price index of the average selling price of newly built buildings in a city and year. We collect the housing selling area and total sales data for each city-year during

---

9 The above specification has no further control of city-level variables except housing price and city-fixed effects. However, the estimation results are consistent if we further control for certain city dimension heterogeneities, such as the GDP growth rate.

10 Using the debt to capital ratio resulted in consistent results.

11 The same method is also used by Liu and Lu (2015).

2003-2007, and then we calculate the average selling price per square meter. The series are deflated with the corresponding provincial CPI index (a city-level CPI index is not publicly available). We then normalize the deflated housing price using the year 2003 as the base year to get the housing price index used in the study (2003=100).

Table 2.3: Variable definitions and summary statistics

Variable	Definition	Obs.	Mean	Std. dev.	P25	Median	P75
Firm level							
Investment	Net change in investment on fixed assets normalized by capital stock (%)	307467	16.48	18.18	4.67	9.05	21.84
Debt	Total debts normalized by total assets (%)	307462	56.71	26.04	38.32	58.00	75.46
Finance	Interests and other finance cost divided by income (%)	307467	1.12	1.66	0.05	0.56	1.54
Labor	Wage and welfare expenditure divided by value added (%)	307467	37.57	38.26	14.63	28.36	48.14
ROA	Earnings before interest and taxes depreciation and amortization normalized by firm assets at the beginning of the year (%)	307467	14.67	18.42	4.56	9.78	18.48
Cash	(Current asset- Current debt)/Total Assets (%)	307467	7.94	28.59	-9.22	7.88	26.13
Industry level							
Industrial median ROA	The median values of ROA at 3-Digit industry level in each year	749	10.45	4.89	8.51	9.88	11.38
Industrial median Cash	The median values of CASH at 3-Digit industry level in each year	749	7.74	5.51	4.84	8.06	10.90
EG index (3-Digit industry, city as the unit)	Agglomeration index (at 3-Digit industry level) calculated using the method developed by Ellison and Glaeser (1997)	749	0.02	0.02	0.01	0.01	0.03
City level							
Ln(HP)	Log of average selling prices of new built commercial building (2003=100)	1085	4.80	0.23	4.61	4.77	4.96
Interest rate	interest rate (% >5 year)	5	6.30	0.64	5.82	6.12	6.46
Instrumental variables							
Share unsuitable land	area with a slope of larger than 15 degrees and the water bodies/jurisdictional size	217	0.22	0.17	0.09	0.20	0.33
Ln(land_GDP)	Log of (land granting area/previous year GDP)	1085	-0.08	0.77	-0.58	-0.02	0.43

Note: Firm- and industry- level data are from Annual Survey of Industrial Firms (ASIF). City-level data is obtained from the China Statistical Yearbook for Regional Economy. Details are explained in Section 2.3.

Firm-level variables of interest are constructed in the following ways. The ratio of earnings before tax and interests to total asset (ROA) is calculated as the sum of the following four items—total profits, sales tax, value added tax and interests in the end of each year—divided by the total assets at the beginning of each year. Similarly, the ratio of cash flow to total assets (Cash) is calculated as the ratio of cash (current asset minus current debt) to the beginning period total assets. These two variables are lagged one year and are expected to be positively correlated with the dependent variable. Firm ownership is defined according to the registration type of a firm, with a dummy variable SOE indicating state-owned enterprise.

We use a range of proxies at the industry level. First, we include the degree of industrial agglomeration (EG index). We choose city as the spatial unit of measuring agglomeration. Industries are defined at the 3-digit level of China Industry Code (CIC) industries. Then, we apply the method developed by Ellison and Glaeser (1997) to calculate the agglomeration index for each 3-digit industry (for more details, see Lu et al, 2012). Second, the median values of ROA and CASH for each 3-digit industry in each year are included to control for heterogeneity in the development circles of different industries.

2.4.2 Identification

It is important to note that housing price may in fact be an endogenous variable. If there are omitted factors that affect both the investment ratio of manufacturing firms and the housing price, or if there exists reverse causality between manufacturing investment and housing prices, the OLS estimates will be inconsistent[12]. The well-accepted approach to isolate the exogenous variation in local real estate prices, used by Mian and Sufi (2010), Chaney et al (2012), and Cvijanovic (2014), is using the land supply elasticity (sometimes interacted with the interest rate or the local trend of real estate prices to allow over-year variations) as an instrument for local housing prices.

In this paper, we propose two instrumental variables. The first one follows the same spirit of using the land supply elasticity to instrument housing prices. As Saiz (2010) finds, land supply elasticity is primarily determined by geographic and regulation constraints on real estate, so we construct a variable measuring the percentage of land in a city unsuitable for real estate development (see also Kung and

[12] Wu et al (2015) point out that it is less likely that there exist omitted variables that can push up housing price but discourage firm investment. The OLS estimates thus should be downward biased.

Chen, 2014). Areas unsuitable for houses consist of land with a slope larger than 15 degrees (based upon architectural safety standards) and bodies of water.[13] We then interact the variable with nationwide movements in the real interest rate because the mortgage rate is an important component of the costs of owning properties and thus affects housing demand.[14] The underlying logic is that cities with less developable land usually see larger changes in housing prices for the same magnitude of demand shocks, e.g., caused by national interest rate changes.

The second instrument is based on the institutional arrangement of China's land supply for urban development. Residential land, including other uses of land, is primarily supplied by local government. We use the amount of urban land transferred via the public land granting market as the other source of potentially exogenous variation in housing prices. We adjust the amount of land granted by the local government by the intensity of local economic activity, as measured by last year's GDP. Cities with relatively scarce land supply, as proxied by land granting area divided by last year's GDP, are likely to witness more growth in housing prices. This gives the following first stage regression model:

$$\ln HP_{c,t} = a_c + d_t + \mathrm{b} * S_c * IR_t + c * \ln(land\_gdp_{ct}) \quad + \omega_{c,t} \ , \tag{4}$$

where a_c and d_t are the full sets of city- and year-fixed effects. $S_c * IR_t$ denotes the interaction between the share of unsuitable land and long term interest rate. $land\_gdp_{c,t}$ is the land-granting area adjusted by local GDP in the previous year. The coefficient b is expected to be positive, whereas parameter c is expected to be negative. There might be reasons to suspect that land area granted by government is endogenous. We therefore proceed in a more cautious manner, sometimes excluding the land-granting area instrument but using the well adopted one only.

---

13 We thank Jianghao Wang for kindly providing us with the data. The detailed procedures to construct the measure can be found in Kung and Chen (2014).

14 We use the benchmark long-term lending rate (above 5-year) as the measure of the mortgage rate. When it changed one or more times in a given year, the interest rate for that year is calculated as the effective days weighted interest rate. The data are from the website of the Peoples Bank of China (http://www.pbc.gov.cn/). http://www.pbc.gov.cn/publish/zhengcehuobisi/631/2015/20150706093731182169554/20150706093731182169554_.html

2.5 Empirical results

To partially avoid selection issues caused by firm entry and exit, we begin by using the balanced panel data to estimate Equations (1) and (2). The estimation results are presented in Section 2.5.1. We then test the heterogeneous and long-term effects of housing price in Sections 2.5.2 and 2.5.3. In Section 2.5.4, we use the unbalanced panel data to examine the robustness of our results when more new firms are included. Last, we investigate the potential mechanism driving the relationship between housing prices and firm investment, if there is one.

2.5.1 Baseline results

We start with estimating the investment equations (1) with the simple ordinary least squares estimator (OLS). Table 2.4 reports the results. In column 1, we include only housing prices and the firm- and year- fixed effects as the explanatory variables. Different from the previous studies, which usually find a positive correlation between housing prices and investment, our estimate of housing prices is significantly negative.

In the following column, we check how the correlation may change when adding other control variables. Therefore, in column 2 we include other firm- and industry- level variables. The coefficient estimate of housing prices ($\ln HP$) remains negative and significant at the 1% confidence level. The coefficient on the log of housing price is now -3.438. Given the average annual growth of housing prices is approximately 10%, this implies that the annual change in housing prices will reduce the firm investment ratio by approximately 0.3438 percentage points. If the growth continues for 5 years (as seen in Table 2.2), the firm investment ratio will thus drop by 1.74 percentage points, a magnitude of economic importance but reasonable.

The estimated coefficients for ROA and CASH are both positive and significant, indicating that profitability and internal cash flows have positive effects on investment. These firm-level determinants are consistent with previous investment literature. The results also show that the product between the initial values of CASH (but not ROA) and the log of housing price also plays a role. The significant positive estimate of the interaction term between CASH and housing prices suggest that firms with more liquidity may buy more properties and benefit from the collateral effect. If the positive effect from CASH dominates the main effect, the CASH variable should be larger than 33 (=3.438/0.077), which is at least the top 25 percentile firms according to the statistics in Table 2.2 (p75 of CASH=26.13). Therefore, for most of the firms in the period studied, their investment ratios are negatively correlated with housing prices.

Table 2.4: Firm investment and housing price: baseline results (OLS)

	(1)	(2)	(3)	(4)
	OLS Full	OLS Full	OLS No BTSC	Weighted OLS Full
Ln(HP)	-11.807***	-3.438***	-4.588***	-2.543***
	(0.583)	(1.317)	(1.183)	(0.925)
ROA		0.089***	0.090***	0.095***
		(0.006)	(0.007)	(0.006)
Cash		0.077***	0.079***	0.072***
		(0.004)	(0.004)	(0.004)
Initial ROA X Ln(HP)		0.015	0.021	0.053*
		(0.030)	(0.033)	(0.030)
Initial CASH X Ln(HP)		0.103***	0.107***	0.077***
		(0.011)	(0.011)	(0.011)
Industrial controls	No	Yes	Yes	Yes
Observations	307,467	307,467	265,229	307,467
R-squared	0.026	0.073	0.069	0.059
Number of firm	64,640	64,640	55,837	64,640

Note: The full sample refers to the balanced firm panel, which has observations for each year of the study period of 2003-2007. The data used in columns (3) are the sub-samples that exclude firms from four municipalities, Beijing, Tianjin, Shanghai and Chongqing (BTSC). The weights equal the inverse of the square root of the number of firms in each city when the weighted OLS method applied. All regressions control for the median values of ROA and CASH at the 3-Digit industry level in each year, with firm- and year-fixed effect. City-fixed effects are actually absorbed by the firm-fixed effects because the firms in our data set are single plant firms. The standard errors of all regression are clustered at the city level as our explanatory variables include the city-level housing price. Significance: *** p<0.01, ** p<0.05, * p<0.1

In column 3, we further verify the results by excluding firms from Beijing, Tianjin, Shanghai, and Chongqing, which are much larger in size and/or experienced much faster growth of housing prices. The negative effects of housing prices become even more pronounced. One possible explanation is that firms sorted into the large cities were more competitive and could adapt more easily to production cost growth.

Lastly, in column 4, we address the concern about whether the varying city coverage may create biases because the number of firms is uneven across cities. To this end, we run weighted OLS estimations using the same specifications as in columns 2. The weights equal the inverse of the square root of the number of firms in the sample for each city. The weighting scheme thus reduces the dominance of cities with more firms in the estimation results[15]. The coefficient estimate for housing price becomes a bit smaller compared with the estimate in column 2 but remains significant.

The results in Table 2.4 should be taken with caution as the estimated coefficients may suffer from various endogeneity biases. We address the issue in Table 2.5 with the instrumental strategy mentioned earlier. In the first two columns of panel A, we suppress the interaction terms between initial values of ROA/CASH with housing price. This yields a simpler model with only one endogenous variable. Meanwhile, we try to be conservative in column 1 by only using the conventional instrument, i.e., the product between the share of unsuitable land for housing construction and national movement in the long-term interest rate. The positive effect of this variable suggests that housing prices increase more as the share of unsuitable land becomes larger, given the size of change in demand, similar to Kung and Chen (2014). The first stage F statistics and its p-values also show no evidence that the instrument is weak. The coefficient estimate of housing prices now becomes -10.543. Equivalently, the investment ratio deceases by 1.0543 percentage points if housing prices grow 10% (i.e., $\Delta \ln HP$=0.1). Bear in mind that the distribution of investment rate is highly right-skewed. One per cent decrease in investment rate will account for approximately 21% of firm investment for firms at the 25 percentile (=1/4.67).

In column 2 we add the other instrument, the log of the land supply to GDP ratio. As expected, binding the supply of land relative to economic activities leads to a higher level of housing prices, ceteris paribus. The F-statistics and Hansen-J test

[15] There is no ex ante reason for either weighed or unweighted estimation to be preferred if we do not know how firms sort across locations.

also provide confidence that the model is neither weakly nor overly identified. The new estimates of the coefficient on housing price are still significantly negative. In column 3, we introduce the products between initial values of ROA and CASH and the log of housing price[16]. This new regression shows very consistent results. Housing price continues to show negative effects on the investment ratio. In panel B, the same procedures are replicated by using the weighted OLS estimators. Weighting has little impact on the results.

Table 2.5: Firm investment and housing price: baseline results (OLS-IV)

	Panel A: IV			Panel B: IV-WGT		
	(1)	(2)	(3)	(4)	(5)	(6)
Ln(HP)	-10.543*	-12.039***	-13.514***	-15.138**	-12.350***	-13.848***
	(5.695)	(4.028)	(4.002)	(6.496)	(4.398)	(4.310)
ROA	0.084***	0.083***	0.084***	0.088***	0.089***	0.090***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Cash	0.071***	0.071***	0.077***	0.068***	0.068***	0.073***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Initial ROA/CASH X Ln(HP)	No	No	Yes	No	No	Yes
First stage Instruments						
Share unsuitable land X Interest rate	0.281***	0.290***	0.261***	0.206***	0.206***	0.169***
	(0.066)	(0.073)	(0.074)	(0.054)	(0.053)	(0.056)
Ln(Land supply/GDP)		-0.033**	-0.033**		-0.032***	-0.033***
		(0.014)	(0.015)		(0.010)	(0.011)
Observations	307,467	307,467	307,467	307,467	307,467	307,467
R-squared	0.069	0.068	0.069	0.054	0.055	0.053
Number of firm	64,640	64,640	64,640	64,640	64,640	64,640
F-stat first stage	13.911	11.731	7.461	9.754	8.583	4.474
p-value F-stat	0.000	0.000	0.000	0.002	0.000	0.000
Hansen J-test		0.574	0.479		0.638	0.663

Note: Initial ROA/CASH X Ln(HP) refers to the interaction terms between housing price and initial values of ROA and cash. They are instrumented when included. For example, the interaction term between ROA in 2003 and housing price is instrumented by, e.g., ROA in 2003 X Share unsuitable land and ROA in 2003 X Ln(Land supply/GDP). All regressions control for the median values of ROA, CASH at the 3-Digit industry level in each year, with firm- and year-fixed effect. The standard errors of all regression are clustered at the city. Significance: *** p<0.01, ** p<0.05, * p<0.1

In sum, our baseline results suggest there was a significant negative effect of housing prices on manufacturing investment in China during the period of

[16] They are also instrumented with the products between the corresponding initials and two instrumental variables.

2003-2007. The magnitude of the impact is also economically large. The OLS estimation results imply that the average coefficient estimate of housing price (log) is approximately -3.438, whereas the IV estimate is roughly as large as -13.514. Given that the full sample standard deviations of the housing price (log) and investment ratio are 0.23 and 18.18, respectively, a one standard deviation change in the log of housing price (i.e., housing prices increase by 23%) can explain 4 to 17 percent of the variation in the investment ratio[17].

2.5.2 Heterogeneity

We next investigate the role of firm and sectoral characteristics. To examine the differential effects of housing price among different types of firms and sectors, we include the following characteristics: firm size, firm ownership, the input-output linkage degree between 3-digit industries and the real estate sector, and the labor intensity of industries. Table 2.6 displays the results.

2.5.2.1 Firm characteristics and heterogeneous responses

Panel A of Table 2.6 shows the differences of responses with respect to two important firm features. In columns (1) and (2), we find small firms may benefit from growth in housing prices. The coefficient estimates of the interaction term between small firms (dummy=1) and housing prices is positive, though it is insignificant in the OLS regression. However, the magnitude of the effect is negligible as the point estimates in both regressions are much smaller than the point estimates of the log of housing prices. This finding could be related to diverse explanations on firm size and collateral effect. For example, on the one hand, small firms are usually credit constrained and need little start-up capital which makes them more likely gain from the collateral effect. However, small firms’ profitability may also be more sensitive to, e.g., the growth in wages pushed by housing price appreciation.

In columns 3 and 4, the SOE dummy are interacted with the log of housing price and are included in regressions. The estimates are sensitive to our choice of

---

[17] 0.23*(-3.438)/18.18 =4% ; 0.23*(-13.514)/18.18=17%

estimating methods. The OLS results imply no significant heterogeneous effects between SOEs and non-SOEs, whereas the 2SLS instrumental variable regression predicts a positive effect for SOEs. A safe prediction therefore would be non-SOEs at least did not invest more than SOEs when real estate prices went up. This finding confronts the impression that non-SOEs are theoretically credit constrained. In that sense, non-SOEs should absorb more collateral effects. Our explanation is that the financial market in China is highly incomplete. The disadvantage of private firms is more than a lack of pledgeable assets (see also Wu et al, 2015).

2.5.2.2 Sectoral characteristics and heterogeneous responses

In panel B of Table 2.6, two industrial features are considered: five quintiles of industrial linkage degree with the real estate sector and five quintiles of sectoral labor intensity. For the first one, we draw the input flow coefficients between each 3-digit industrial sectors to the real estate sector from the input-output matrix of China. The quintile of industries is based on this coefficient. The labor intensity quintiles are based on the median values of the labor-to-capital ratio for all of the 3-digits industry-level during our study period.

In Columns 5 and 6 of Table 2.6, we estimate the heterogeneous effect with respect to sectoral linkage with the real estate industry using both OLS and 2SLS estimators. Overall, an industry that is highly linked with the real estate sectors has larger coefficient estimates. However, the effect is not quite significant. Only the top quintile significantly increases investment compared to other categories (according to result in column 6). The same regressions are also performed for the labor intensity categories. We find that the higher the labor intensity of an industry, the more investment is reduced. The set of quintile dummies are all significantly negative and larger in the upper categories of labor intensity (columns 7 and 8).

Overall, the positive spillovers from the industrial linkage channel are not statistically strong. This may because the sectoral spillovers may be beyond the geographic unit we study, i.e., city. However, we find robust evidence that the more labor-intensive sectors reduce more investment. The evidence is consistent with the hypothesis that housing price appreciation raises local factor costs and thus discourages the willingness to invest.

Table 2.6: Firm investment and housing price: firm and sector heterogeneity

	Panel A: Firm characteristics				Panel B: Sectoral characteristics			
	OLS	OLS-IV	OLS	OLS-IV	OLS	OLS-IV	OLS	OLS-IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Ln(HP)	-3.460***	-13.210***	-3.435***	-13.619***	-3.502***	-13.726***	-2.792**	-12.833***
	(1.320)	(4.078)	(1.317)	(4.000)	(1.334)	(3.994)	(1.265)	(4.055)
Small firm X Ln(HP)	0.102	0.094*						
	(0.062)	(0.057)						
SOE firm X Ln(HP)			-0.061	0.207**				
			(0.083)	(0.094)				
Q2 of Real estate link X Ln(HP)					0.059	0.233		
					(0.112)	(0.160)		
Q3 of Real estate link X Ln(HP)					-0.129	-0.155		
					(0.133)	(0.244)		
Q4 of Real estate link X Ln(HP)					0.140	0.187		
					(0.132)	(0.162)		
Q5 of Real estate link X Ln(HP)					0.166	0.435**		
					(0.140)	(0.187)		
Q2 of Ln(Labor_K) X Ln(HP)							-0.466***	-0.535***
							(0.112)	(0.184)
Q3 of Ln(Labor_K) X Ln(HP)							-0.598***	-0.595***
							(0.146)	(0.192)
Q4 of Ln(Labor_K) X Ln(HP)							-0.849***	-1.006***
							(0.156)	(0.192)
Q5 of Ln(Labor_K) X Ln(HP)							-0.928***	-0.976***
							(0.197)	(0.245)
Instrumental variables	No	Yes	No	Yes	No	Yes	No	Yes
Observations	307,467	307,467	307,467	307,467	307,467	307,467	307,467	307,467
R-squared	0.073	0.069	0.073	0.069	0.073	0.069	0.073	0.069
Number of firm	64,640	64,640	64,640	64,640	64,640	64,640	64,640	64,640
F-stat first stage		5.576		5.632		4.488		3.884
p-value F-stat		0.000		0.000		0.000		0.000
Hansen J-test		0.335		0.131		0.442		0.328

Note: All regressions control for ROA, CASH, and initial values of ROA(CASH) interacted with housing price (log), three industry-level variables (median values of ROA, CSAH for each industry-year, and the EG index), together with the full sets of firm- and year-fixed effects. Small firms are defined as the firms which own assets less or equal to the median amount of assets in their city in a year; while others are defined as the large firms. Quintiles of real estate link are based on the input flow coefficients between each 3-digit industry and real estate sector. The data are compiled from 2007 Input-Output Tables of China. Quintiles of labor intensity are calculated based the median values of the log labor-to-capital ratio during the study period of 2003 to 2007. The instruments used are the products between the share of unsuitable land and the interest rate and log of land supply adjusted by local GDP (lagged one year). Variables interacted with housing price are also instrumented by the corresponding products with two instrumental variables. The standard errors of all regression are clustered at the city. Significance: *** p<0.01, ** p<0.05, * p<0.1

2.5.3 Temporary versus sustaining effect

To detect whether the correlation between housing prices and manufacturing investment exists in the long run, we do the first difference regressions using only data at the starting and ending years. Panels A and B of Table 2.7 apply OLS and 2SLS estimators, respectively. Because the regressions look only at the cross firm-city variations, the IV regressions include only the share of unsuitable land as the instruments[18].

The OLS results in Panel A (column 1) suggest the negative effects of housing prices on investment still hold. The estimated coefficient is now much larger than the estimates for the annual panel data, indicating that firms gradually reduce their manufacturing investment, possibly due to the positive capital adjusting cost.

Columns 2-5 investigate the heterogeneous effects in the long term. We still did not find significant collateral effects for small firms. In contrast, the OLS estimates in column 3 imply that SOE firms may increase their investment in the long term. The net effect for SOEs now is 2.690 (i.e., =7.836-5.146). Regarding the sectoral heterogeneity, we still find that labor-intensive industries invest less, whereas the industries with higher degrees of linkage with real estate sectors invest more. This confirms that the labor cost channel is an important channel through which housing prices discourage firm investment.

Panel B in Table 2.7 presents the IV regression results. The new estimates are much larger than the long-term OLS estimates and the IV estimates of the annual panel. To gauge its economic significance, supposing housing prices grew 50% during 2003-2007 (roughly 10% per year), the investment ratio will decrease by 11.5 percentage points (= -23*0.5), which is about the range between p75 and p50 in our sample. Firm and sector characteristics play mediating roles similar to those described in the previous results.

[18] Although the land supply variable has between-group variations, its effect is theoretically present only in the short term.

Table 2.7: Firm investment and housing price: long term effect

	(1)	(2)	(3)	(4)	(5)
Panel A: OLS					
ln(HP)	-4.960**	-5.182**	-5.146**	-5.255**	-4.440*
	(2.299)	(2.382)	(2.289)	(2.336)	(2.393)
Small firm X ln(HP)		0.699			
		(0.625)			
SOE firm X ln(HP)			7.836***		
			(0.811)		
High linked industries X ln(HP)				0.639	
				(0.412)	
High labor intensity X ln(HP)					-1.312**
					(0.610)
ROA	0.107***	0.106***	0.107***	0.107***	0.107***
	(0.011)	(0.010)	(0.010)	(0.011)	(0.010)
CASH	0.056***	0.056***	0.057***	0.056***	0.057***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Observations	48,909	48,909	48,909	48,909	48,909
R-squared	0.030	0.030	0.031	0.030	0.030
Panel B OLS-IV					
ln(HP)	-22.791***	-23.042***	-22.742***	-23.180***	-21.999***
	(8.131)	(8.107)	(8.129)	(8.159)	(8.003)
Small firm X ln(HP)		1.207*			
		(0.716)			
SOE firm X ln(HP)			5.071**		
			(2.190)		
High linked industries X ln(HP)				0.984**	
				(0.471)	
High labor intensity X ln(HP)					-1.816**
					(0.789)
ROA	0.095***	0.094***	0.095***	0.095***	0.096***
	(0.013)	(0.013)	(0.014)	(0.013)	(0.013)
CASH	0.056***	0.056***	0.057***	0.056***	0.057***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
First Stage					
Share unsuitable land	0.331***	0.334***	0.339***	0.334***	0.350***
	(0.114)	(0.115)	(0.113)	(0.115)	(0.115)
F stat first stage	8.396	4.198	4.199	4.198	4.196
p-value F-stat	0.004	0.016	0.016	0.016	0.016
Observations	48,909	48,909	48,909	48,909	48,909
R-squared	0.010	0.011	0.012	0.010	0.011

Note: All regressions control for initial values of ROA and CASH. The standard errors of all regression are clustered at the city. Significance: *** p<0.01, ** p<0.05, * p<0.1

2.5.4 Endogenous entry and exit

The above analysis focused on the pre-existing continuing firms. The choice allows us to have more within-firm variations, which helps in the identification. However, concern arises whether the negative effects of housing price growth on firm investment exist for new emerging firms or exiting firms. The challenge to answering that question is that firms can observe housing price levels and their changes, which can determine the location choice of firm investment. Therefore, in the investment regressions with new emerging firms, our instrumental strategy may not help. Nevertheless, we repeat our baseline analysis using the unbalanced panel. Columns 1 - 4 in Table 2.8 contain the results. In the first two columns, we consider the panel of firms with at least 4 - year observations. The results are close to those with balanced panels. A similar pattern holds when we use all the firms with at least 3 - year observations; however, the IV regressions report insignificant estimates of housing price in column 4[19].

Table 2.8: Endogenous entry and exit of firms and housing price

	(1) > = 4 OLS	(2) > = 4 OLSIV	(3) > = 3 OLS	(4) > = 3 OLSIV
Ln(HP)	-3.303***	-7.869*	-2.862**	-3.550
	(1.167)	(4.508)	(1.164)	(4.914)
ROA	0.085***	0.083***	0.081***	0.080***
	(0.007)	(0.007)	(0.007)	(0.007)
Cash	0.095***	0.095***	0.100***	0.100***
	(0.004)	(0.004)	(0.004)	(0.004)
Observations	527,662	527,662	648,395	643,530
R-squared	0.063	0.062	0.060	0.060
Number of firm	134,037	134,037	187,012	182,147
F-stat first stage		5.039		4.598
p-value F-stat		0.000		0.000
Hansen J test		0.309		0.097

Note: Regressions of columns 1-4 also control for ROA, CASH, and initial values of ROA(CASH) interacted with housing price (log), three industry-level variables (median values of ROA, CSAH for each industry-year, and the EG index), together with the full sets of firm- and year-fixed effects.

2.5.5 Mechanism

Thus far, the paper has established that housing prices have robust negative impacts on industrial firms' investment, whereas most studies find that the appreciation of housing prices relaxes credit constraints and therefore stimulates fixed asset investment. In this section, we explore the mechanisms that explain the negative association that we find in our study. To do so, we estimate econometric models specified by Equation (3).

In panel A of Table2.9, we apply the OLS estimator to investigate how housing

[19] The other choice for understanding the impact of housing price on the entry and exit of local firms is to estimate a growth equation of the numbers of entry/exit firms, instead of the investment equation we used in this paper. However, the ASIF data set have only information on firms above 500 million Yuan in sales. We are thus not able to have the number of new firms below this threshold, unless census data are available. As far as we know, the NBS has done three economy censuses that cover all economic units in both secondary and services sectors (2004, 2008 and 2013). The data would facilitate a better understanding of this topic. We leave it for the future studies.

prices impact the three intermediating variables. We find that housing prices have significant positive impacts on finance and labor cost but not the debt-to-asset ratio. This evidence is consistent with the Dutch disease type of effect, whereby prices of production factors are driven up because of housing price appreciation. This will, at least in the short-term, discourage investment. In panel B, we continue to find that finance cost is significant and positively correlated with housing price. However, the labor cost share becomes insignificant. Firms may reduce the number of workers as labor costs increase. Overall, we take these results as suggesting that housing prices might push up the price of production costs, including credit and labor. The composition effect of housing prices on firms' choice of production factors, especially how the labor market responds to housing price changes, is worthy of its own new study.

Table 2.9: Firm investment and housing price: mechanisms

	Panel A: OLS			Panel B: OLS-IV		
	(1)	(2)	(3)	(4)	(5)	(6)
	Debt	Finance	Labor	Debt	Finance	Labor
Ln(HP)	0.823	0.487***	5.603**	9.455***	1.500***	1.933
	(0.934)	(0.124)	(2.657)	(3.109)	(0.514)	(7.120)
ROA	-0.084***	-0.004***	-0.058***	-0.080***	-0.003***	-0.057***
	(0.007)	(0.000)	(0.008)	(0.007)	(0.000)	(0.008)
CASH	-0.062***	-0.002***	0.021***	-0.061***	-0.002***	0.021***
	(0.007)	(0.000)	(0.005)	(0.007)	(0.000)	(0.005)
F-stat first stage				7.458	7.458	7.458
p-value F-stat				0.000	0.000	0.000
Hansen J test				0.224	0.067	0.190
Number of firm	64,640	64,640	64,640	64,640	64,640	64,640
Observations	307,462	307,462	307,462	307,462	307,462	307,462
R-squared	0.033	0.012	0.004	0.029	0.001	0.004

Note: The sample used for regression is the balanced firm panel during the study period 2003-2007. All regressions control for initial values of ROA (CASH) interacted with housing price (log), three industry-level variables (median values of ROA, CSAH for each industry-year, and the EG index), together with full sets of firm- and year-fixed effects. Instrumental variables used in columns 4- 6 are both the product between share of unsuitable land and interest rate and the log of land supply divided by local GDP (lag one year). The standard errors of all regressions are clustered at the city level. Significance: *** p<0.01, ** p<0.05, * p<0.1

2.6 Conclusions

This paper investigates the effects of booming housing prices in China on firm investment using a large panel data set of Chinese manufacturing firms over the period 2003-2007. We find significant negative effects of local housing prices on firm investment. The results contradict with the well documented positive relationship between housing prices and firm investment — the so-called collateral effect. Further exploration reveals that such negative effects of housing prices on investment are stronger for firms in the labor-intensive sectors, suggesting real estate price appreciations may push up production costs and thus discourage investment incentives — a mechanism similar to the Dutch disease effect.

Though we use instrumental variables to address the endogeneity issue of real estate prices. The results should be interpreted cautiously. First, our estimates do not convey general equilibrium effects. Specifically, housing price may have positive spillovers to other cities. A negative relationship between housing prices and firm investment therefore does not strictly imply a negative relationship between the two at the national level. Second, our data set does not allow us separate machinery equipment investments with other investments, especially land or real estate holding investments. However, given our estimated effects of housing price on firm investment are negative, there would be even larger negative effects of housing price appreciations on firm investment when investment on land or real estate holdings are excluded from the total investment.

CHAPTER 3

Exposure to the Real Estate Boom and Intra-industry Resource Allocation Efficiency: Industrial Level Evidence from China

Abstract

This study provides industrial-level evidence on the impacts of the real estate boom on sorting of upstream manufacturing firms in China during 2000-2007. Using intrinsic linkage with the real estate sector and the sudden boom of the real estate market after 2003, we apply the difference-in-differences approach to identify the intra-industry resource allocation effects of the real estate boom. Our results show that industries intrinsically highly linked with real estate sectors experienced increasing disparities in firm total factor productivity (TFP), suggesting sorting-in or expansion of less-efficient firms in more-exposed industries. The results are robust to a few alternative model specifications.

Keywords: Real estate boom, Firm sorting, TFP dispersion, Misallocation

JEL Classification L11 L60

This chapter is about to be submitted to Journal of Comparative Economics, as Wu, Y., N. Heerink, L. Yu. 'Exposure to the Real Estate Boom and Intra-industry Resource Allocation Efficiency: Industrial Level Evidence from China.'

Land and Labor in Urbanizing China

3.1 Introduction

Many economies have experienced dramatic fluctuations in real estate markets (Shiller, 2015). Recent examples include property price booms and boosts in Japan (1980s—1990s) and in the US (1990s —early 2000s) and the steep housing price increase in China since the early 2000s. Among these cases, fluctuations of real estate market are often found have detrimental impacts on the economy at large. The world economy today, to some extent, is still coping with the aftermath of the 2007-09 financial crisis, which was initially triggered by the home mortgage crisis in the US.

Recent research to understand the comprehensive and intensive impacts of real estate market on an economy has been growing. Over the past years a few studies find that the large impacts of real estate market is due to its collateral function – as one important form of collateral for both households and firms, real estate value gains (loss) can relax (strengthen) financial constraints for (household) consumption or (firm) investment (see Chaney et al., 2012; Gan, 2007; etc.). However, there are potentially other channels that link real estate market with the real economy. As can be observed from input-output tables, the real estate industry has broad and deep linkages with the others sectors (Chan et al., 2016). Rigorous investigations of the industrial linkage effects are still scarce. Moreover, few studies have examined an important question: how a real estate boom reshapes intra-industry resource reallocation (in upstream manufacturing industries).

This study posits that real estate booms lead to sharp increases in product demand in linked upstream industries; increasing demand can cause the decrease in competition, and therefore allows the entry and expansion of less-efficient firms. As consequences, resource allocation efficiency will be harmed by real estate booms. This effect follows the spirit of the early literature on intra-industry resource reallocation (see the discussions in Section 3.3.3), which find that tougher competition in an industry will yield greater similarity of firm performance (e.g., larger market share for more efficient firms). Conversely, a decrease in inter-firm competition can allow the expansion and entry of low-productivity firms and, consequently, increased dispersion of firm performance.

We empirically test, for any given industry, whether more-efficient firms benefit more (or less) from diffusions of the real estate market using a rich firm- and industry-level dataset for China during 2000-2007. Specifically, we use the dispersion of firm total factor productivity (TFP) within an industry, which is linear with resource allocation efficiency, as the main outcome of interest (see Hsieh and Klenow, 2009). We apply a difference-in-differences (DID) framework: using the degree of intrinsic linkage with the real estate sector, generated from the sector input-output (I-O) table, as the measure of industries' exposure to the real estate market; meanwhile, using the sudden booms in the real estate market after the year 2003 as a turning point. We examine whether the interaction terms between industrial linkage and the timing of the sharp growth in real estate prices help explain industrial-level TFP distribution.

Our empirical results show that the interaction term between industrial linkage and the timing of the sharp growth in real estate prices has a positive effect on TFP dispersion — exposure to the real estate boom increases TFP inequality and thus reduces intra-industry resource allocation efficiency. These results are also robust to different model specifications. Our interpretation is twofold. First, real estate expansion fuels demand in linked industries; this rising demand, in turn, drives up market prices and thus profits in the linked industries. Higher returns further lower entry standards and therefore provide opportunities for marginal investors—who are generally less efficient—to enter the markets. Second, by contrast, land, capital, labor and other production costs increase due to diffusion of the rapid growth in housing prices. Rising costs force less-efficient firms, especially in industries with fewer linkages with the real estate sector, to exit.

Our study is closely related to recent work on the spillover effects of the real estate market boom in China on the rest of the economy, particularly manufacturing industries. These studies have revealed various effects that the real estate sector can have on China's manufacturing sectors. For example, Chen and Wen (2014), using a calibrated model, show that the fast growth of housing prices in China incentivizes firms to divert resources away from productive capital accumulation, such as research and development, and can hamper the growth of TFP. These predictions have gained support from recent studies. For instance, Rong et al. (2016) document evidence that housing price growth in China reduces manufacturing firms' innovation investment. In addition, Chen et al. (2015) provide firm-level evidence that housing price appreciation in China crowded out firms' fixed capital investment related to their main business. In contrast to the above literature, we add the intra-industry allocation effect, an overlooked mechanism, to the spillovers that the real estate boom can have on China's manufacturing sectors. More generally, our study aligns with recent studies seeking to incorporate the heterogeneity of individuals/entities into economic theory (e.g., Banerjee and Duflo, 2005; Melitz, 2003).

The rest of this paper is organized as follows. Section 3.2 provides background on the real estate market in China. Section 3.3 discusses the theoretical foundations. The empirical framework and results are presented in Sections 3.4 and 3.5, respectively. Concluding remarks are then offered in Section 3.6.

3.2 Background

The recent real estate boom in urban China follows two major policy reforms by the central government in 1998 and 2003, which first transformed the employer provision of housing to the current market-based housing demand-supply system and then accelerated its development, respectively. In this section, we briefly review the series of policy reforms in China's housing market and provide descriptive evidence that these policies have dramatically altered the subsequent development path of the urban housing market.

3.2.1 Housing policy reforms in China

The housing market is relatively new in China. When the Communist Party came into power in 1949, the Chinese government nationalized urban land and suppressed home-ownership—with the exception of pre-existing home owners, who were allowed to retain property rights to their residences. New urban residents all rented houses, at nominal rates, built and distributed by their working units (usually state-owned enterprises or government agencies). Housing during that era was thus part of social welfare attached to jobs; its costs mainly covered by government budgetary revenues.

This system, however, met difficulties in practice, including the financial burden of housing construction, poor management, and corruption in housing allocation. Policy experiments were then implemented throughout the 1980s to address these problems, including increasing housing rental rates and privatizing public houses (see Wang and Murie, 1996, for more details). In July 1994, after a long time of policy trials and revisions, the State Council decided to privatize houses: tenants in employer-provided houses were given the opportunity to buy their current homes. [20] Consequently, many (state-sector) employees began to own property, and housing demand emerged (Wang, 2012). Following the successful introduction of housing ownership, the State Council, with its 23rd Decree in 1998, further prohibited employers from building and allocating properties to employees.[21] Instead, houses must be bought or rented from the housing market. The employer-provided in-kind housing benefits were transformed into monetary housing subsidies. Due to this policy change, commercial housing producers no longer faced competition from employer-provided houses. A market-based housing supply and demand system were finally laid out in 1998 (Wu et al., 2012).

The 1998 reform set the foundations of the commercial housing market in urban

20 The Decree is titled 'The Decision on Deepening the Urban Housing Reform'

21 The Decree is titled 'A Notification on Further Deepening the Reform of the Urban Housing System and Accelerating Housing Construction'

China, but the urban land market did not become apparent until the year 2002. Urban land for development was usually acquired from local governments (the de facto land owners) through negotiation (Wu et al., 2015). This policy has been criticized for its vulnerability to bribery and corruption and resulted in land prices below market value. In May 2002, the Ministry of Land and Resources required all residential and commercial land parcel leaseholds to be sold via public auctions after July 2002. The policy, due to rigidity, was not fully executed until early 2004, when the Ministry of Land and Resources and the Ministry of Supervision jointly issued 'Notification on the continuation of business land use rights for transferring auction listing of law enforcement work notice' (Cai et al., 2013). As documented in many studies, the enforcement of land auctions pushed up land prices as well as housing prices dramatically.

In March 2003, five years after the 1998 reform, the State Council issued 'A notification on promoting the sustained and healthy development of the real estate market [No.18]'. The significance of the decree related to the declaration by China's top-level government that the 'real estate sector is one pillar industry of the national economy — for its wide inter-industry linkages and strong driving effects on other sectors. ' The notion of 'pillar industry' was interpreted as endorsement of the real estate sector by the government; public confidence in real estate development thus strengthened, even though the intention of the No.18 decree was to slow housing price growth, mainly in large cities.

3.2.2 Impacts of housing reform: descriptive evidence

Housing policy reforms have had profound impacts on China's real estate market. The annual amount of housing space supplied by the private market increased from approximately 25 million square meters in the mid-1980s to approximately 500 million square meters in 2007 (Wu et al., 2012). The appreciation of housing prices was also remarkable. Figure 3.1 shows the evolution of housing prices since the 1998 housing reform. In 1999, the national average housing price was approximately 2000 Yuan per square meter; in 2007, it was close to 4000, thus nearly doubling in 8 years.

Despite concerns about the reliability of Chinese official housing-price data, this overall trend of housing prices is consistent with public impression. The year 2003, most notably, marked an important turning point in the housing market. Before 2003, housing prices overall grew steadily, less than 5% per year; by contrast, the annual growth rate became much higher after 2003. According to an estimate from Wu et al. (2012) based on data from Beijing, land prices, which account for 30-40% of housing prices, increased by 200% between 2003 and 2008. Evidently, the sharp change in housing price growth is related to the land auction reform around 2003.

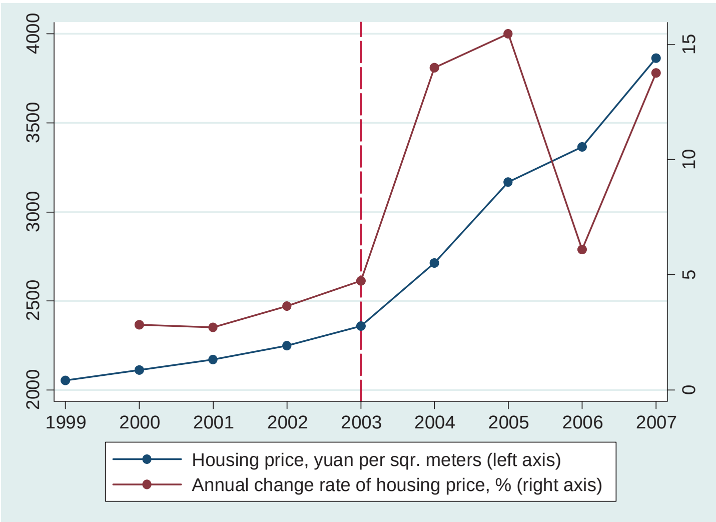


Figure 3.1: National average housing price and its change rate in China (1999-2007)

Note: The national average housing price is calculated as the ratio of total sales of commercial buildings to the total area. Source: China Statistical Yearbook (various years) and authors' calculations.

In addition, as shown in Figure 3.2, the operation performance of the real estate sector started to increase after the 1998 housing reform. By 2004, profit margins in the sector surpassed those in the manufacturing sector, which hovered around 7%. In subsequent years, the profit premium in the real estate sector continued to increase, until 2010. The stagnation in industrial sector performance, given the fast economic growth (and thus demand side) in the same period, is surprising and implies there might be little productivity gains from the supply side.

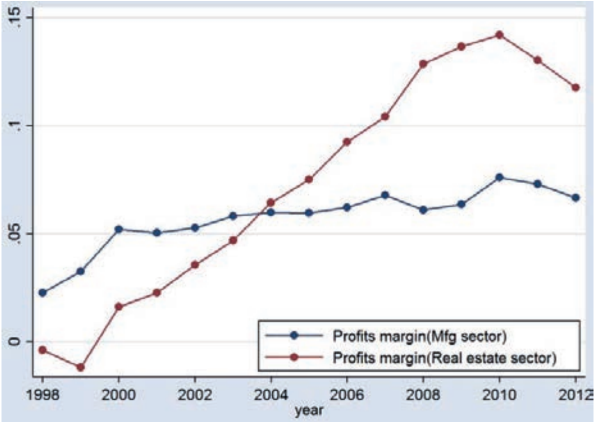


Figure 3.2: The divergence of profit margins (industrial sector vs. real estate sector: 1998-2012)

Note: profit margin = profits/revenues Source: China Statistical Yearbook (various years) and authors' calculations.

3.3 Theoretical motivation and hypothesis

Real estate, in the modern economy, plays vital roles beyond purely shelter. It is widely agreed that the real estate sector has both financial and industrial linkages with the rest of the economy that operate through financial networks and the real economy, respectively (see Chan et al., 2016). In this section, we outline important findings in the literature on the impacts of the real estate market on the economy at large. We then focus on the possible intra-industry resource allocation effects of the real estate boom.

3.3.1 Financial and industrial linkage effects

Real estate has broad, intense linkages in an economy via both financial and real channels. The prior literature has mostly focused on financial linkage, and rigorous analyses on industrial linkage effects are scarce. Real estate has financial linkages because housing assets are important collateral for issuing loans in modern economies (Gan, 2007). When there is financial friction, appreciation of housing values can help relax financial constraints, thus increasing access to credit to households and firms. Many studies have consequently documented positive relationships between housing prices and firm investment. Similar positive effects are also often observed for household consumption (for more details on this aspect, see Black et al., 1996; Chaney et al., 2012; Mian and Sufi, 2011, etc.).[22] By contrast, industrial linkages have received limited attention. In a qualification of the importance of the real estate sector in the Chinese economy, Chan et al. (2016) incorporate both financial and industrial linkages. In particular, they use total input coefficients in input-output table data to determine the real spillovers from the real estate sector.

3.3.2 Real industrial linkages and exposure to the real estate boom

Many sectors are suppliers and/or consumers of the real estate sector, and they differ in terms of the magnitudes of inter-dependence. The classic method for exploring real linkages between industries is input-output (I-O) tables, which describe the real flows of goods and services between sectors. These tables usually contain two indexes of the relative interdependence between sectors: the direct input coefficient and total input coefficients. The former measures only the direct inputs but does not consider, e.g., the production process of intermediate inputs. The total input coefficients, by contrast, consider both direct and induced input-output relationships. Specifically, the total input coefficients illustrate how much outputs in each sector are used (both directly and indirectly) for one unit of output in the sector of interest (Chan et al., 2016). Therefore, the total input coefficients are more appropriate for measurements of true inter-industry dependence. Moreover, the real estate sector, according to the present industrial classification in China's statistics, refers to housing services such as real estate development, administration and agencies, whereas the construction of houses belongs to the construction industry. The use of total input coefficients can also minimize impact from the exclusion of

[22] This conclusion that there is positive effect of real estate price increases on investment or consumption is subject to debate: there is also evidence against this conclusion (for details see Wu et al., 2015).

construction stages.[23]

Figure 3.3, which uses data from China's input-output table 2007, depicts the linkages between the industrial sector and real estate sector.[24] The length of the horizontal bar represents the total input coefficients from the real estate sector: the inputs used to produce one unit of output in the real estate sector. As shown in the figure, the chemical and metal industries have been most closely linked with the real estate sector, followed by the fuel, electric and electrical equipment sectors.

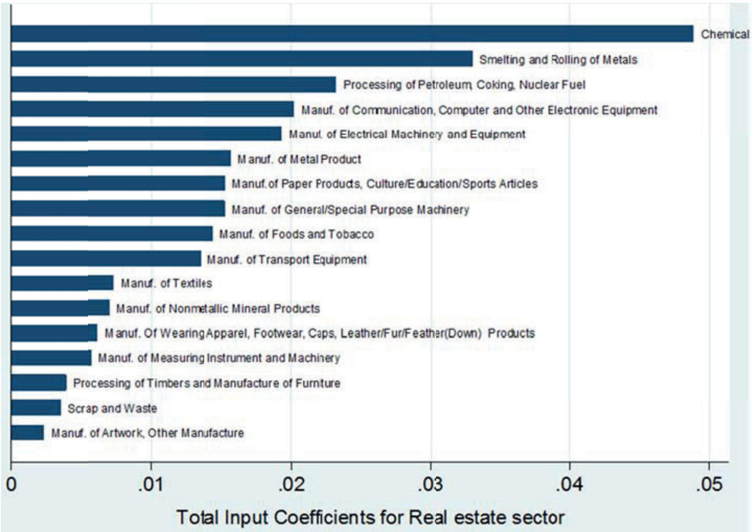


Figure 3.3: The linkages between the industrial sector and real estate sector
Source: Input-Output Tables of China (2007)

One may wonder if the real linkages, in terms of total input coefficients, vary over time. In fact, the inter-sector differences in their linkages with a sector, according to Chan et al. (2016) and the two recent input-output tables for China, are relatively small, which is plausible as the technology of production in a particular sector remains roughly unchanged, particularly in the short term. Consequently, it is not

[23] Some studies have attempted to aggregate the total input coefficient from both the real estate and construction sectors (Chan et al., 2016). This method, however, adds the impact from other non-real estate construction activities, such as infrastructure, that play more significant roles in the construction sectors. We opt for a conservative definition of the real estate sector but a comprehensive multiplier—the total input coefficients.

[24] One alternative input-output table for the year 2002 was not considered; its industry classifications are more aggregate and thus allow much less variation in inter-industry linkages.

problematic to presume that the real exposure to real estate development, as measured by total input coefficients, is intrinsic, i.e., exogenous in an econometric sense.

3.3.3 Real estate shock and intra-industry resource reallocation

Intra-industry resource reallocation arises from the finding of firm heterogeneity, usually in terms of productivity differences. For example, Banerjee and Duflo (2005) note the important disadvantages of new classical economics, in which the assumption of a representative agent in the economy ignores the fact that individuals are so different from many perspectives. The seminal work of Melitz (2003) incorporates firm heterogeneity in trade models. His study reveals an important firm-sorting mechanism related to trade openness: trade openness helps the most-efficient firms expand, whereas less-efficient firms opt to produce in a home market, and the least-efficient firms exit altogether. Moreover, Alfaro et al. (2008), Restuccia and Rogerson (2008), and Bartelsman et al. (2013) show evidence that differences in the allocation of resources across heterogeneous plants are a significant determinant of cross-country differences in income per worker.

The inclusion of heterogeneity in economic models has trigged numerous studies and yield two important insights (see Bernard et al., 2011 and Syverson, 2011 for good reviews). First, the dispersion of firm heterogeneity, especially in a narrowly defined industry or a small scale space can carry information about resource allocation efficiency. For example, Hsieh and Klenow (2009) use the dispersion of TFP (in narrowly defined industries) to qualify the extent of misallocation in China and India. Also, Ryzhenkov (2016) applies the same method to Ukraine to qualify resource misallocation.

Second, given the importance of firm heterogeneity, many studies have attempted to understand the causes of firm heterogeneity. Syverson (2004) argues that product substitutability in an industry determines the within-industry competition — industries with low transportation costs or fewer advertising expenditures (i.e., high substitutability) exhibit lower productivity dispersion. Laeven and Woodruff (2007) and Alfaro and Chari (2014) link firm size distribution to a variety of institutional features; both studies demonstrate that institutional quality has heterogeneous impacts on firm size (e.g., different percentiles have different impacts). Both Lu and Yu (2015) and Edmond et al. (2015) investigate the relationship between trade openness and industrial-level firm markup dispersion. These authors find that pro-competitive gains from international trade by reducing markup dispersions. In general, the literature has consistently indicated that as conditions approach the first best world (e.g., absence of uncertainty, market frictions, bad institution and/or policy distortions), firm performance becomes more similar, and vice versa.

The fundamental proposition of this study—stronger exposure to real estate demand shock causes firm TFP dispersion to increases– follows similar logic: the sudden increase in housing demand and housing prices drive up demand for and, consequently, the prices of products in highly linked industries; the quasi windfall gains allow small and less-talented entrepreneurs to sort into highly linked industries. Firm size and total factor productivity (TFP) would become more dispersed in industries that are more exposed to the real estate sectors. For less-linked industries, housing prices push up labor wages, among other production costs, and less-efficient firms may not able to afford the changes and are thus forced to exit. Accordingly, these industries will see a decrease in TFP dispersion.

3.4 Empirical strategy

3.4.1 Model specification

The strategy employed in this study exploits the timing of the real estate boom and the differential degree of industrial linkage between 3-digit manufacturing industries and the real estate industry. The specification of the regression equation is thus as follows:

$$y_{i,t} = \alpha + \beta \times Link_i \times Post03_t + \boldsymbol{X'}_{i,t} \times r + \mu_i + \sigma_t + \varepsilon_{i,t} \ , \qquad (1)$$

where i denotes industry and t represents year. $y_{i,t}$ denotes the sectoral-level productivity-dispersion measure, such as the log of the Theil index of TFP. $Link_i$ represents the linkage between an industry and the real estate sector based on the input-output table. $\boldsymbol{X}_{i,t}$ is a vector of time-varying industrial characteristics. The variables μ_i and σ_t capture industry and year fixed effects, respectively. $\varepsilon_{i,t}$ is an error term. The inclusion of industry fixed effects absorbs unobserved heterogeneity in the industry-specific determinants of productivity, whereas the year dummies control for macroeconomic shocks common to all industries. The parameter of interest is β, which measures the impact of linkage with the real estate sector on economic growth in the post-relative to the pre-real estate boom period.

The above specification is essentially a difference-in-differences (DID) approach. A valid identification of β, however, requires that no other changes occurred near 2003 that impacted both the dependent variable and the proxy of industrial linkage between the manufacturing sectors and real estate sector or, formally, $Cov\ (Link_i \times Post03_t, \varepsilon_{i,t}) = 0$. This requirement is plausible in our study because industrial linkages can be viewed as an intrinsic characteristic, particularly in the short term. The determination of the true identity of β is related to the intention-to-treat effect. For example, firms may observe the evolution of their production cost. They prepare for this by, for example, adjusting their production technology to minimize impacts. Conversely, firms may also prepare for upcoming investment opportunities in real estate-linked industries. Hence, the extent to which the real estate boom was unexpected and the rapidity with which it spread to other industries are of importance for our DID method. We propose that most Chinese firms did not prepare well for the arrival of real estate expansion. Although firms intended to prepare, their production technology was not easily changed within a short period of time. In addition, it may take additional time for new firms to emerge in linked industries. Nonetheless, to verify our DID identification, we estimate a more general form of equation (1) as follows:

$$y_{i,t} = \alpha + \sum_{t \in D^{pre}} \beta_t \times d_t \times Link_i + \sum_{t \in D^{post}} \beta_t \times d_t \times Link_i + \boldsymbol{X'}_{i,t} \times r + \mu_i + \sigma_t + \varepsilon_{i,t} \quad , \qquad (2)$$

where $\{d_t\}_{t \in D^{pre}}$ and $\{d_t\}_{t \in D^{post}}$ represent the years before and after the real estate boom, respectively. If our DID approach is valid, the coefficient set β_t before the year 2003 will not be significantly different from zero, whereas the counterparts after 2003 will be (jointly) significant if the real estate boom affected productivity distribution[25].

Another concern is that the indicator before or after the real estate boom ($Post03_t$) is only an approximate measure. In comparison with the discrete measure for the real estate boom, we opt for national real estate investment or housing prices as continuous proxies for the real estate boom. The model can then be replaced by the following specification:

$$y_{i,t} = \alpha + \beta \times Link_i \times RE_{i,t} + \boldsymbol{X'}_{\boldsymbol{i,t}} \times \gamma + \mu_i + \sigma_t + \varepsilon_{i,t} \quad , \qquad (3)$$

where $RE_{i,t}$ is a continuous variable that measures the real estate boom (i.e., national real estate investment) for each year (log). However, real estate investments may be endogenous and bias the coefficient estimation of β. For instance, the price of manufactured goods (manifested through productivity measures) may have a feedback effect on real estate investment. Despite these shortcomings, we examine the specification as a robustness check.

3.4.2 Data and variable definitions

3.4.2.1 Data sources

Our raw data are from the Annual Survey of Industrial Firms (ASIF) conducted by the National Bureau of Statistics of China (NBS). These data are the most comprehensive firm-level data in China and are used widely in academic studies. As we need to match these data with the total input coefficients data, we aggregate the NBS firm-level data at the 3-digit China Industry Classification (CIC) industry level. The study period spans from 2000 to 2007, before the onset of the global recession in

[25] In the long term, productivity distribution can be independent of inter-industrial linkage with the real estate sector, i.e., the treatment effects disappear.

2008. Consequently, annual panel data at the 3-digit industry levels are applied in the empirical analysis—approximately 150 industries for 8 years.

3.4.2.2 Variable definitions

Measuring productivity and its dispersion

Total factor productivity, by definition, is the optimal choice when measuring economic efficiency or resource misallocation. We follow the methodology of Levinsohn and Petrin (2003) to get our estimates of firm level TFP. [26] Specifically, the estimation is based on a C-D form production function. The elasticities of inputs are industry specific. The inputs and outputs are revenue based, and deflated. To gauge TFP dispersion, we calculate the Theil index of firm-level TFP. Alternative dispersion measures such as the inter-quantile range are also considered. For example, we use the distance between the 75th and 25th percentile standardized by the median industry values for each industry-year (as in Syverson, 2004).

Qualifying real estate exposure

We determine the input flow coefficients between each of the 3-digit industry sectors and the real estate sector from the input-output matrix for China. China published input-output tables for the years 1992, 1997, 2002 and 2007. However, only the 2007 table contains comprehensive information at relatively disaggregated levels (3-digit industry).

Control variables

Following Syverson (2004) and Lu and Yu (2015), we include five time-varying industrial-level variables as our control variables. Specifically, we use total exports in an industry divided by total outputs to measure export intensity. According to the theoretical mechanism highlighted in Melitz (2003), access to foreign markets can help efficient firms expand and thus reduce dispersion. We include the degree of industrial agglomeration (EG index). To calculate the index, we denote city as the

[26] There are several methods for estimating TFP. A detailed discussion of TFP estimation can be found in recent studies such as Syverson (2011).

spatial unit to measure agglomeration. This variable captures the degree of geographical concentration for an industry. The number of firms and the mean value of fixed assets are controlled as proxies for barriers to entry (it is more difficult for industries with more firms and higher fixed-asset requirements to enter).

3.5 Results

3.5.1 Main results

Table 3.1 reports the results from the fixed-effects estimation of equation 1. Column 1 shows the regressions of the Theil index of TFP (log) on the product of the industry-specific linkage degree with the boom-period indicator (Post 2003 =1 if during 2004-2007, otherwise 0). The estimated coefficient on the term equals 0.061 and is significant at the 5% level. Column 2 includes the time-varying industrial variables. The interaction term remains significant and positive. Because we standardized the linkage variable (mean of link=0, standard deviation of link=1), the estimate 0.072 (column 2) can be interpreted as follows: ceteris paribus, an increase of one standard deviation in the degree of linkage with real estate sectors will increase the Theil index of TFP by 7.2%; in other words, the inequality of firm TFP will be 7.2% greater compared with that of an industry with a degree of linkage one standard deviation lower.

In columns 3 and 4, we add the product of the linkage with the real estate sector and a year dummy 2003 to determine if the linkage degree plays a role in determining the outcome variables. This is the minimum requirement (i.e., no pre-boom effect of industrial linkage with the real estate sector) for the reliability of the identification in columns 1 and 2. The term indicates an insignificant effect for both columns, providing confidence that the DID approach is appropriate. The baseline results thus indicate that exposure to the real estate boom indeed causes larger inter-firm inequality of TFP (within narrowly defined industries).

Table 3.1: Real estate boom and productivity dispersion: baseline results

	(1)	(2)	(3)	(4)
	DV: Theil Index of TFP (Log)			
Link X Post 2003	0.061**	0.072***	0.064**	0.076***
	(0.027)	(0.026)	(0.029)	(0.028)
Link X 2003			0.010	0.015
			(0.020)	(0.019)
Control variables	No	Yes	No	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Industry fixed effects	Yes	Yes	Yes	Yes
Observations	1,120	1,120	1,120	1,120
R-squared	0.261	0.275	0.261	0.276
Number of sic3	143	143	143	143

Note: Standard errors are robust and clustered at the 3-digit industry level. The list of control variables includes the mean value of fixed assets (log); export share; SOE Share; EG Index, and number of firms (log). *** p<0.01, ** p<0.05, * p<0.1

For the control variables (not reported in the table, available upon request), two time-varying industrial variables, export/output and the EG index, are also determined to be significant. The negative coefficient estimates of export intensity imply that access to international markets facilitates within-industry competition, consistent with the new-new trade theories for heterogeneous firms (Melitz, 2003). The EG index captures the geographical agglomeration of an industry, and its significant and positive coefficient indicates that geography-specific rent results in a wide dispersion of inter-firm TFP. The number of firms (log) and the mean value of firm fixed assets are insignificant. However, as previously mentioned, the estimates for control variables focus more on correlation than on causation.

3.5.2 Identification verification

As discussed previously, the minimum requirement of equation 1 is that there is no significant relationship between industrial linkage with the real estate sector and productivity distribution before the housing construction boom. By contrast, the post-boom estimates should be significant, but the extent depends on the timing of the real estate boom, how quickly the treatment effect appeared, and how long it lasted. We empirically test this concern by estimating equation 2: regressing the Theil index of TFP (log) on a full set of interaction terms between year dummies and the industrial linkage measure.

Panels A and B in Table 3.2 report results using a full or short panel of data, respectively. The use of a short panel (i.e., the time period 2002-2005) focuses on a short period near the starting point of the housing boom. We thus avoid the impacts of confounding year-specific shocks and/or minimize other endogenous responses of firms. The results in columns 1 through 4 consistently indicate that a positive correlation between exposure to real estate and the dispersion of TFP began when the real estate boom started in 2005—approximately 1-2 years after the treatment year.

Table 3.2: Real estate boom and productivity dispersion: identification verification

	(1)	(2)	(3)	(4)
	DV: Theil Index of TFP (Log)			
	Panel A: Data used 2000-2007		Panel B: Data used 2001-2005	
Link X 2000	0.006	0.000		
	(0.028)	(0.028)		
Link X 2001	-0.036	-0.045*	-0.036	-0.045*
	(0.024)	(0.024)	(0.024)	(0.025)
Link X 2002	0.000	0.000	0.001	0.000
	(0.018)	(0.018)	(0.017)	(0.018)
Link X 2004	0.017	0.024	0.017	0.024
	(0.026)	(0.025)	(0.026)	(0.025)
Link X 2005	0.055*	0.062**	0.055*	0.063**
	(0.029)	(0.027)	(0.029)	(0.027)
Link X 2006	0.065**	0.071**		
	(0.033)	(0.031)		
Link X 2007	0.077**	0.087***		
	(0.031)	(0.030)		
Control variables	No	Yes	No	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Industry fixed effects	Yes	Yes	Yes	Yes
Observations	1,120	1,120	700	700
R-squared	0.265	0.280	0.222	0.238
Number of sic3	143	143	143	143

Note: Standard errors are robust and clustered at the 3-digit industrial level. The list of control variables includes the mean values of fixed assets (log); export share; SOE Share; EG Index, and number of firms (log).
Reference category = Link X2003.
*** p<0.01, ** p<0.05, * p<0.1

3.5.3 Robustness: alternative outcome variables

In Table 3.3, we use several alternative outcome variables to check the robustness of the baseline results. First, we replace the dependent variable with the inter-quartile range of TFP (standardized by TFP of the median firms), which is easier to interpret. The results in column 1 indicate that one standard deviation increase in the linkage degree with the real estate sector increases the inter-quartile TFP difference by 0.2%;

the effect is significant at the 1% level. In column 2, we choose the specifications with all of the year dummies. The linkage variables, however, become insignificant.

Table 3.3 Robustness: alternative outcome variables

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	TFP(#75-#25)/#50		Theil Index of Y (Log)		Theil Index of L (Log)	
Link X Post2003	0.002*		0.025***		-0.003	
	(0.001)		(0.009)		(0.012)	
Link X 2000		-0.001		-0.010		0.012
		(0.002)		(0.014)		(0.014)
Link X 2001		-0.002		-0.011		0.022**
		(0.002)		(0.013)		(0.010)
Link X 2002		0.000		-0.025**		0.005
		(0.001)		(0.011)		(0.012)
Link X 2004		0.000		0.021**		0.013
		(0.001)		(0.009)		(0.011)
Link X 2005		0.002		0.010		0.006
		(0.001)		(0.009)		(0.008)
Link X 2006		0.003		0.006		0.002
		(0.002)		(0.011)		(0.010)
Link X 2007		0.002		0.018*		0.006
		(0.002)		(0.011)		(0.010)
Control variables	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,200	1,200	1,168	1,168	1,168	1,168
R-squared	0.059	0.063	0.189	0.192	0.193	0.196
Number of sic3	150	150	148	148	148	148

Note: Standard errors are robust and clustered at the 3-digit industrial level. The list of control variables includes the mean value of fixed assets (log); export share; SOE Share; EG Index, and number of firms (log).
Reference category = Link X2003.
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Second, we use firm size, firm output or employment as an alternative proxy for productivity to partially resolve the concerns over the robustness of TFP estimations. There are caveats in using output or employee as the proxy for productivity, despite evidence that firm size and efficiency are closely associated (Bartelsman et al., 2013). For instance, firm output may be related to firms' pricing behavior; the number of employees can be associated with the firms' financial constraints and/or ownership types (in China). Nevertheless, firm size is a reasonable metric for productivity and is commonly used when the information to derive TFP is insufficient. The corresponding results suggest that the use of total outputs to calculate dispersion measures yield evidence consistent with that obtained using TFP (columns 3 and 4). However, the dispersion in employment size is not significantly impacted by real estate exposure. In addition to the potential causes mentioned above, it is more likely due to the asymmetric labor-market effects of a demand shock: it is easier for more highly linked firms to hire more workers than for less-linked industries to lay off workers (especially in the short term).

3.5.4 Robustness: alternative definitions of exposure to the real estate boom

Table 3.4 tests the robustness of the baseline results using different definitions of exposure to the real estate boom, as specified in equation (3). Column 1 uses the product of industrial linkage and real estate investment at the national level (log). Column 2 replaces real estate investment with housing sales (log), and column 3 uses the ratio of real estate investment to GDP. We use models with several definitions of firm-performance disparity, which are provided in each panel of Table 3.4. We note consistent results with the DID approach. The results, therefore, are robust to alternative definitions of exposure to the real estate boom.

Table 3.4 Robustness: alternative definitions of real estate exposure

	(1)	(2)	(3)	(4)
Panel A DV = Theil Index of TFP (Log)	RE=LN(REI)	RE=LN(Housing Sales)	RE=LN(Housing Price)	RE=%(REI/GDP)
Link X RE	0.040***	0.042***	0.043***	0.025***
	(0.014)	(0.014)	(0.014)	(0.009)
Panel B DV = TFP(#75-#25)/#50				
Link X RE	0.001*	0.001*	0.001*	0.001*
	(0.001)	(0.001)	(0.001)	(0.001)
Panel C DV=Theil Index of Y (Log)				
Link X RE	0.012**	0.012**	0.012**	0.008**
	(0.005)	(0.005)	(0.005)	(0.003)
Panel D DV=Theil Index of L (Log)				
Link X RE	-0.004	-0.004	-0.003	-0.002
	(0.007)	(0.007)	(0.006)	(0.004)

Note: All regressions include all time-varying industrial characteristics, as well as the full set of industry and year fixed effects. Standard errors are robust and clustered at the 3-digit industrial level.
*** p<0.01, ** p<0.05, * p<0.1

3.6 Conclusions

This study investigates how the real estate boom alters competition between firms of different productivity levels. Our results show that industries that are intrinsically highly linked with real estate sectors experience increasing heterogeneity in firm TFP after a real estate boom—suggesting sorting-in or expansion of less-efficient firms in sectors which are highly linked with real estate sectors. Our identification strategy is also novel. We make use of both differences in the intrinsic linkages with the real estate sector and the sudden boom of the real estate market after 2003 in a classic DID framework. The most important implication of our study is that the real estate boom reduces inter-firm competition pressure and thus reduces resource allocation efficiency.

CHAPTER 4

Foreign Direct Investment, Fiscal Decentralization and Land Conflicts

Abstract

Land disputes have been an important risk to social stability in China since the turn of the century. This paper uses provincial data on illegal land uses during the period 1999–2010 as a proxy for the intensity of land conflicts to investigate the effects of foreign direct investment (FDI) and fiscal decentralization on jurisdictional land conflicts. The results show that the FDI growth rate has a positive and significant impact on the growth rate of illegal land use when there is a high degree of fiscal decentralization. We thus provide evidence supporting the hypothesis that regional competition for FDI, as shaped by fiscal decentralization, tends to raise conflicts over land in China.

Keywords: FDI, Fiscal decentralization, Land conflicts, China

JEL classification：R11, R52, H70

This chapter has been published as Wu, Y. & Heerink, N. (2015). Foreign direct investment, fiscal decentralization and land conflicts in China. China Economic Review, 38, 92-107.

4.1 Introduction

Land has been a major source of conflict in China over the last decade. As Cao et al. (2008) report, there were approximately 17,900 cases of “massive rural incidents” in the first nine months of 2006. Among these, land acquisition-triggered conflicts accounted for approximately 80% of all incidents. In the recent work of Kung and Chen (2014), land revenue windfalls are found to have led to increases in corruption, as measured by the size of the bureaucracy and the amount of administrative expenditure. Cui et al. (2015) investigate the impacts of land requisition on farmers’ perceptions of the government. Unsurprisingly, they find that land taking increases political distrust between villagers and local officials.

Although it is widely recognized that land disputes increasingly contribute to social instability in China, few studies attempt to investigate the underlying causes of land conflicts. A lack of reliable data on land conflicts is probably the main reason for this research gap. In this study, we propose to use available data on illegal land uses as a proxy for land conflicts. Data on illegal land uses are collected and reported periodically by the Ministry of Land and Resources. The ministry and its sub-branches play a major role in supervising land use behavior and the implementation of land use administration laws in contemporary China.

We focus on two mechanisms identified in the literature as major causes of land conflicts in China. Both mechanisms originate in China’s fiscal decentralization policy. One explanation comes from the literature on regional competition in China. It emphasizes that land is an important instrument of local governments in attracting foreign direct investment (FDI) and stimulating local economic growth. The other is that land can be leased out in the urban land market to the real estate sector and other tertiary-sector businesses for prices that far exceed those paid to farmers for expropriated land; this so-called land finance serves as an important source of revenues for local governments (Cao et al. 2008; Tao et al. 2010; Tan et al. 2011).

Guided by the literature, we examine the effects of FDI, land finance and fiscal decentralization on land conflicts using a provincial-level panel data set for the 1999–2010 period. The results suggest that: (1) Regional competition for FDI has positive (short-term) effects on the growth of illegal land use when there is a relatively high degree of fiscal decentralization. (2) Land finance, as a share of local fiscal budgetary revenues, does not significantly affect land conflicts.

Our findings are consistent with the literature on FDI and its impacts on host economies. For example, Robertson and Teitelbaum (2011) show that FDI leads to increases in labor protests, especially when the host economy has inadequate labor rights protections. Our sub-national study similarly finds that the combination of outside capital and deficient protection of land rights can generate serious predation of farmers' land. The results confirm local conditions are determinative in the host country.

The remainder of the paper is organized as follows. Section4. 2 discusses the use of illegal land use as a proxy of land conflicts in China. Section 4.3 presents our analytical framework and main hypothesis. Section 4.4 describes the empirical strategies and the data set. The main results and sensitivity checks are then discussed in Section 4.5. Section 4.6 concludes.

4.2 Land conflicts in China

Many studies contend that land-triggered conflicts have become a primary cause of social unrest. For example, Lu et al. (2012) report that approximately half of China's mass incidents (protests involving more than 100 participants) in 2012 were caused by land disputes.[27] The rise of land disputes is also reflected in the growth of land-related petition letters (Meligrana et al. 2011; Pils 2005). Estimates based on petition letters received by the People's Congress of China in 2012 show that over 50% of petitions concerned land issues such as unfair compensation following land acquisition.[28] These findings, however, are mainly based on incomplete statistics on mass incidents and petition letters at the national level. Data of this type are only released occasionally in government documents and do not allow for quantitative analysis of the variations in land disputes across jurisdictions and over time. To gauge land conflicts and systematically investigate their causes, we propose to use data on illegal land use areas

[27] Mass incidents refer to "planned or impromptu gathering[s] that [form] because of internal contradictions", including mass public speeches, physical conflicts, the airing of grievances, or other forms of group behavior that may disrupt social stability. A typical example of a mass incident trigged by land is the Wukan case (see *http://en.wikipedia.org/wiki/Wukan_protests*).

[28] Petitioning is a commonly used instrument by Chinese citizens to lodge complaints when their rights are infringed upon as a result of the abuse of power by authorities, public institutions, enterprises, or civil groups.

as a proxy for the local intensity of land disputes. We introduce the data and explain their suitability in this section.

4.2.1 Illegal land use as a proxy

Land use in China is mainly governed by the Land Administration Law (LAL) that was first promulgated in 1986. It was significantly revised in 1998 to meet the challenges posed by the rapid loss of farmland and fast increase in government requisitions of rural land. The revised law defines rules for land requisition and leasing and thereby governs the allocation of land between the rural (mainly agricultural) and urban sectors (Tan et al., 2011; Cao et al., 2008).

During the period 1999–2008, on average, more than 171,000 hectares of farmland per year were expropriated for (mainly urban) construction purposes (Table 4.1). After expropriated farmland is transferred from rural collectives to urban governments, it can be leased through the urban land market via the public land leasing system, with a maximum lease term of 40 years for commercial use, 50 years for industrial use, and 70 years for residential use. Government revenues obtained from land leasing through the urban land market reached more than 2,746 billion Yuan in 2010 (Table 4.1).

Table 4.1: Land leasing, farmland losses and illegal land use in China, 1999-2010

Year	Farmland Loss for Construction Uses (Hectares)	Land Leasing Area (a) (Hectares)	Land Leasing Revenues (b) (10^6 Yuan)	Land Leasing Price (a/b) (10^6 Yuan/Hectare)	Illegal Land Use Area (Hectares)
1999	134679.86	45390.68	51432.95	1.13	17809.44
2000	163258.89	48633.22	59558.48	1.22	22647.50
2001	163653.95	90394.12	129588.96	1.43	20219.57
2002	196499.60	124229.84	241679.25	1.95	19898.33
2003	229105.72	193603.96	542131.13	2.80	35761.85
2004	145089.57	181510.36	641217.60	3.53	36468.18
2005	138694.58	165586.08	588381.71	3.55	32355.46
2006	167368.24	233017.88	807764.47	3.47	61381.65
2007	188285.96	234960.59	1221672.08	5.20	62426.16
2008	191568.19	165859.67	1025979.88	6.19	32719.41
2009	-	220813.90	1717952.56	7.78	27591.24
2010	-	293717.81	2746447.91	9.35	27895.20

Source: MLR (2000-2011)

Following the LAL revision in 1998, the Ministry of Land and Resources and its sub-national branches began to document incidents of illegal land use. MLR (2000 – 2011) distinguishes six categories of illegal land use: illegal transfer, damage to cultivated land, encroachment on land without approval, unlawful approval of land occupancy, granting of land at reduced prices, and others. Among these categories, the encroachment on land without approval accounts for the largest share (Table 4.2). It refers to the occupation of land without legal approval or the occupation of more land than the approved area. As anecdotal evidence illustrates, illegitimate land occupation by land users and illegal approval of land conversion have been important reasons for land disputes among farmers, investors, and local governments.[29] Because the amount of land involved in illegal land use cases has been systematically documented at the province level since 1999, we can use its variation across provinces and over time to examine its causes in detail. In our view, it is the best available proxy for the intensity of land conflicts that can be used for an empirical examination of their causes.[30]

Table 4.2: Shares of illegal land use types in total illegal land use, 1999-2010 (%)

Year	Illegal transfer	Damage to Cultivated land	Encroachment on land without approval	Unlawful approval of land occupancy	Granting of land at reduced prices	Others
1999	13.6	3.7	46.6	1.6	0.3	34.2
2000	11.6	3.5	57.3	10.1	1.0	16.5
2001	10.9	5.1	71.3	3.3	0.0	9.4
2002	7.7	3.6	80.2	0.5	0.0	7.9
2003	6.3	3.1	84.4	1.1	0.1	4.9
2004	5.2	3.1	84.6	1.0	0.8	5.4
2005	3.5	5.1	85.7	0.2	0.0	5.5
2006	3.4	2.8	79.5	8.0	1.5	4.9
2007	3.8	2.2	87.8	1.8	0.0	4.3
2008	3.2	2.1	90.1	1.0	0.0	3.6
2009	1.4	2.7	88.5	2.1	0.0	5.3
2010	1.1	1.6	90.6	2.2	0.0	4.4

Source: Calculated by authors from MLR (2000-2011)

It should be noted that not all illegal land conversion cases are conducted by local officials and that there are also (an unclear number of) cases committed by normal civilians or by firms themselves. Yet, even in such cases, local officials often play an important role. Many illegal cases committed by civilians or firms involve official connivance. To localize investment or to sell land, officials have strong incentives to

[29] Two typical examples of illegal land use can be found at:
http://www.chinadaily.com.cn/china/2011-10/26/content_13977191.htm
http://www.chinadaily.com.cn/bizchina/2010-10/18/content_11423012.htm

[30] The sub-type "damage to cultivated land" mainly refers to houses built on farmland or trees planted on farmland instead of crops, which has little to do with FDI and fiscal decentralization. The total number of illegal land use cases of this type is very small, and its impact on our analysis is therefore likely to be negligible. We decided to focus on the data on illegal land use as a whole, as data collection is considerably less demanding under that approach.

open a window (back-door) for illegal uses.[31] For these reasons, cases committed by civilians or firms have not been excluded from our analysis.

4.2.2 Illegal land uses - trends and spatial distribution

In 1999, nearly 18,000 hectares of land were found to be illegally used. This value peaked in 2006 at more than 62,000 hectares and then declined to less than 28,000 hectares in 2009–2010 (see the last column in Table 4.1). Figure 4.1 shows the spatial distribution of land conflicts, as measured by the average land area involved in illegal use cases during the 1999–2010 period divided by the total land area in a province. The hot spots of land conflicts, represented in dark colors in the figure, are provinces located in the east of China, particularly Shandong, Jiangsu, Beijing and Tianjin, and Hainan in the south. Guangdong in the south and Henan in the center of China also have high levels of illegal land use. By contrast, the western and northern provinces, including Xinjiang, Inner Mongolia, Heilongjiang, Qinghai, Gansu, Sichuan, and Yunnan, have relatively little land involved in illegal uses. Generally, the spatial distribution reflects China's regional economic development.

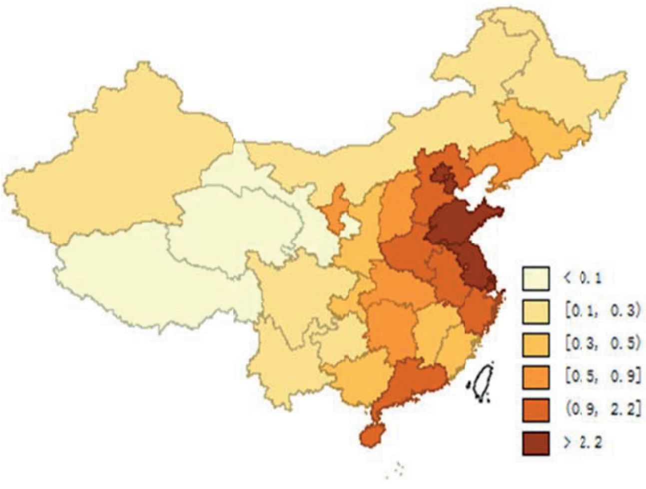


Figure 4.1: Geographical Distribution of Land Conflicts (Illegal land use area per 10, 000 hectares), 1999-2010

Notes: Data on illegal land use area come from China Land and Resources Almanac (MLR, 2000-2011). For each province, the illegal land use areas are first weighted by the jurisdiction area and then averaged over the years.

[31] Additional information (in Chinese) on collusion between government officials and normal civilians or firms can be found on the following websites: *http://news.xinhuanet.com/fortune/2006-06/11/content_4679218.htm* *http://news.xinhuanet.com/house/2007-03/21/content_5875154.htm*

4.3 Framework and hypothesis

This section reviews the available literature on channels through which FDI, land value windfalls and fiscal decentralization are expected to affect land conflicts. Based on the resulting framework, we establish the main hypothesis to be tested.

4.3.1 FDI inflows and land conflicts

A large body of literature examines the effects of FDI on the economic development of the host economies. Recent studies argue that the impacts of FDI are context-specific. The absorptive ability and/or the local (political) environment are found to be crucial in explaining the sizable differences observed in the growth effects of FDI for different countries or regions. For example, several studies show that positive growth effects of FDI are more likely when a host economy has a high level of human capital, a more developed financial sector, and/or greater economic freedom (Borensztein et al., 1998; Bengoa and Sanchez-Robles, 2003; Alfaro et al., 2004). Negative effects of FDI, particularly the "race to the bottom" of labor and environmental standards, are more likely to occur when a recipient economy has a less developed democracy, ill-defined property rights, and/or an underdeveloped market economy (e.g., Lan et al., 2012; Dean et al., 2009; Cole et al., 2006). In addition, FDI can have spillover effects on local governance and social relations. For example, Dang (2013) finds evidence that FDI in Vietnam improved the quality of local governance, while Robertson and Teitelbaum (2011) report that FDI increases labor protests in developing countries with weak labor rights protections.

China has followed a policy of actively attracting FDI since its opening-up and reform policies that began in the late 1970s. In 2011, China received FDI inflows worth 123,985 million USD, equal to 7.5% of global FDI flows in that year and 1.72% of Chinese GDP (UNCTAD, 2012). Accumulated FDI through 2011 is estimated to be equivalent to 711,802 million USD, making China the largest FDI recipient in the world. To host foreign investment, sizable areas of land are needed to construct industrial parks. Between 2003 and 2006, the number of "development zones" and industrial parks in the country nearly doubled (Tao et al., 2010).

The increased demand for land that results from FDI inflows may not result in higher land prices. To attract FDI to their cities or towns, government officials tend to offer low land prices (and other amenities), thereby causing a "race to the bottom" (Tao et al., 2010). However, because attracting FDI increases their promotion opportunities, land serves as a valuable resource for local officials. Inflows of FDI may also have positive spillover effects by creating more and/or better employment opportunities and by improving local governance quality. In such cases, FDI inflows may not cause land conflicts or even reduce their incidence. Overall it is therefore ambiguous how the increased demand for land that results from FDI inflows impacts land disputes.

4.3.2 Land value windfalls and land conflicts

In the literature, the resource (windfall gains) – conflict nexus is a subject of intense debate (see Nillesen and Bulte, 2014). The emerging consensus is that whether resource booms trigger conflicts is conditional on factors such as institutional quality, political regimes and types of resources. An economy can benefit from the windfall gains of resource rents only if appropriate institutions have been installed before resource booms occur (Deaton, 1999; Mehlum et al., 2006; Robinson et al., 2006). Booms in point-source and capital-intensive resources are more likely to trigger conflicts (Ross, 2015).

This study focuses on land windfall gains generated by converting rural land into urban land and the leasing out of urban land to the tertiary sector by local governments. A few recent studies report supportive evidence that the windfall gains in land are the source of many social problems in contemporary China. For example, Kung and Chen (2014) show that land revenue windfalls have led to the rapid expansion of the size of the bureaucracy size and corruption. The findings in Cui et al. (2015) imply that land requisition encourages mistrust between farmers and local governments.

4.3.3 Fiscal decentralization and land conflicts

China has become a highly decentralized economy since its fiscal reforms in the 1980s and 1990s. Many studies analyze the effects of China's fiscal decentralization on the country's economic growth (for a review, see Xu, 2011). One major finding is that fiscal decentralization encourages local governments to invest in public infrastructure to cultivate their tax bases, thereby promoting economic growth. Using variations in ex post fiscal decentralization at the provincial level, fiscal decentralization has been identified as a driving force of economic growth in China (Jin et al., 2005; Lin and Liu, 2000; Qiao et al., 2008). However, the growing importance of local governments due to decentralization has been found to contribute to regional protectionism, inflation, jurisdictional disparity and other problems (Feltenstein and Iwata, 2005; Zhang, 2006). Wang (2013) formalizes the relationship between fiscal decentralization and the incentives for attracting FDI in a political-economy model. The model predicts a non-monotonic effect of fiscal decentralization on FDI inflows. China currently lies in the medium range of decentralization, with local governments being highly motivated to attract FDI.

Recent work on fiscal decentralization notes that land prices in China are used as an important instrument to attract investors (Tao et al., 2010). A common approach is for local authorities to acquire land from farmers and then develop well-equipped industrial units for investors. The land prices of these units are usually very low to reduce the investment costs of investors, especially foreign investors, who substantially enhance the "image" of local development efforts. In many cases, land acquisition and preparation costs in industrial zones may be much higher than leasing revenues, and compensations for farmers may be kept low to balance budgets. Therefore, attracting FDI frequently results in conflicts with farmers (Liu et al., 2008; Tao et al., 2010; Yang and Wang, 2008).

Moreover, fiscal decentralization also increases the pressure on local governments to raise revenues. With insufficient tax revenues, they will rely more on "land finance" to cover their expenditures (Nitikin et al., 2012). The pressure to raise revenues and the incentive to attract FDI thereby lead to aggressive land acquisition and may ultimately result in conflicts with expropriated farmers (Cao et al., 2008).

4.3.4 Main hypothesis

The channels connecting FDI, fiscal decentralization and land conflicts, according to the literature, are summarized in Table 4.3. First, FDI inflows can raise land values and increase the risk of land conflicts. However, the positive effects of FDI on local governance and employment may mitigate the negative effects, making the overall impacts on land conflicts ambiguous. Second, fiscal decentralization places pressure on local governments to raise revenues. When farmers' land rights are not sufficiently protected and local governments have a dominant position in land requisition, fiscal decentralization potentially encourages aggressive land grabs and thus creates land conflicts. Third, the motivation to compete for FDI and economic growth is strongly shaped by ex post fiscal decentralization. The use of land to attract FDI will therefore be more intense in more decentralized regions. As a result, we hypothesize that the interaction between FDI and fiscal decentralization contributes positively to land conflicts.

Table 4.3: Mechanisms connecting FDI, fiscal decentralization and land conflicts

	Foreign Direct Investment	Fiscal Decentralization
Fueling conflicts	Increasing land value, i.e., raising stakes under contest and thus conflicts	Fiscal self-reliance, greedy governments, etc. → race to the bottom in (industrial) land prices
Avoiding conflicts	Better governance; more employment opportunities; higher compensation for farmland	

4.4 Empirical strategy and data

The main aim of our empirical analysis is to examine the (separate and combined) effects of FDI and fiscal decentralization on land conflicts. Our basic econometric strategy is therefore to determine whether the growth of FDI and the degree of local fiscal decentralization can help to explain changes in land conflicts, measured as the growth in illegal land use area. Additionally, we investigate the land windfalls channel. In this section, we discuss the econometric strategy and the data used.

4.4.1 Empirical model specification and estimation

Our basic regression model is as follows

$$\Delta lnLC_{it} = a + \beta 1 \Delta lnFDI_{it} + \beta 2FD_{it} + \beta 3\Delta lnFDI_{it} * FD_{it} + \beta 4\ LF_{it} + \beta 5X_{it} + \lambda_t + \mu_i + \varepsilon_{it} \quad (1)$$

where $\Delta lnLC_{it}$ is the change in the logarithm of illegal land use area in province *i* in year *t*; it can be interpreted as the rate of change in illegal land use area. Following a common approach used in the growth empirics literature (e.g., Durlauf et al., 2005), we specify all explanatory variables as either growth rates or ratios. $\Delta lnFDI_{it}$ refers to the change in the logarithm of FDI inflows divided by population size for province *i* in year *t*. FD_{it} denotes the fiscal decentralization measure. LF_{it} is land finance, i.e., land leasing revenue divided by government budgetary revenue; it can be interpreted as the government's degree of fiscal dependence on land requisition and leasing. X_{it} is a vector of the province-level control variables. λ_t denotes a full set of year dummies to account for time-specific effects. μ_i is the unobserved province-specific effect, which captures time-invariant regional effects such as geographic location, climate and culture. ε_{it} is a random disturbance term. Finally, a and β_1, ... β_5 are the unknown coefficients.

The coefficients β_1 and β_3 are of primary interest in testing our main hypothesis. As explained in Section 4.3.1, the sign of coefficient β_1 could be positive or negative, while the sign of β_3, the coefficient of the interaction term $\Delta lnFDI_{it} * FD_{it}$, is expected to be positive.

Fiscal decentralization affects land conflicts in three different ways, as described in Section 4.3.3: not only through its interaction with FDI but also through land finance (LF) and its impact on officials' incentive to promote economic growth (represented by FD as a stand-alone variable in our model). The coefficient of LF is expected to be positive, while the coefficient of FD could take either sign. The latter coefficient reflects the impact of FD on officials' incentives to promote economic growth and, thereby, on urbanization and industrialization (for given levels of FDI and LF) and also on regional protectionism and other growth-inhibiting policies.

To estimate the model, we used the fixed effects estimator – more precisely, the within estimator – to account for provincial fixed effects. Random effects estimation was not considered for two reasons. One is that the fixed effects estimator is more appropriate than the random effects estimator when unobserved time-invariant characteristics are correlated with the predictors (Wooldridge, 2012). The other advantage of the fixed effects approach is that it can help address the endogeneity issues of regressors if those endogenous regressors are correlated with a time-invariant component of the error (Carmeron and Triveda, 2010). In our case, it is likely that FDI is correlated with, for example, distance to a port (a conventional instrument for FDI), which varies across provinces but usually not over time. The endogeneity driven by it will naturally be solved by using the fixed effects approach. We also include year fixed effects in the model to account for potential endogeneity caused by omitted variables that vary systematically over time and may be correlated with the dependent variable, such as financial crises or global food prices. Yet, even when using province and year fixed effects, there may still exist significant correlation between FDI and region- and time-specific aspects of economic development or other omitted variables. In particular, our main variable of interest (i.e., $\Delta lnFDI_{it}$) may be a proxy for region-specific economic growth. To examine this possibility, we analyzed the correlation matrix of our explanatory variables. The correlation coefficients of the growth rate of FDI per capita and the other explanatory variables in our model range between -0.06 and 0.14. Hence, we conclude that the correlation between FDI and region- and time-specific aspects of economic development or other omitted variables is negligible.

4.4.2 Data sources and variable definitions

The data set used in the empirical analysis covers 30 provinces over the 1999–2010 periods. Tibet was excluded because of missing data. The period and units of observation were chosen based on the availability of data on the dependent variable of interest, i.e., the illegal land use area. Table 4.4 lists the variable definitions and sources. Table 4.5 presents summary statistics.

Table 4.4: Variable Definitions and Sources

Variables	Description	Source
Land conflicts (LC)	Illegal land use area (hectares)	CLRA
FDI per capita (FDI)	Foreign Direct Investment per capita (2005 US$)	CSYRE
Fiscal decentralization1 (FD1)	Budgetary fiscal revenue / budgetary spending	CSY
Fiscal decentralization2 (FD2)	Local fiscal income per capita /(local fiscal income per capita + central fiscal income per capita)	CSY
Land finance	Land leasing revenues / budgetary fiscal income	CLRA& CSY
GDP growth rate	Real GDP growth rate	CSY
Urbanization	Non-agricultural population / regional population	CSY
GDP per capita	GDP per capita (1,000 yuan, 1999 price)	CSY
Gov. consumption ratio	Government consumption/total consumption	CSY
SOE employment share	State owned enterprise employment /total wage employment	CSYRE

Notes: CLRA = China Land and Resources Almanac (MLR, 2000-2011); CSYRE = China Statistical Yearbook for Regional Economy (NBSb, 2000-2011); CSY = China Statistical Yearbook (NBSa, 2000-2011)

Table 4.5: Descriptive statistics

Variable	Observations	Mean	Std. Dev.	Min	Max
Δln (illegal land use area)	330	0.073	0.858	-2.794	4.693
Δln (FDI per capita)	330	0.094	0.213	-1.293	2.017
Fiscal decentralization (FD1)	330	0.516	0.188	0.148	0.951
Fiscal decentralization (FD2)	330	0.441	0.151	0.255	0.863
Land finance	330	0.372	0.285	0.004	1.705
GDP growth rate	330	0.121	0.024	0.054	0.238
ln (GDP per capita)	330	9.475	0.655	7.928	11.120
Urbanization rate	330	0.347	0.160	0.145	0.889
Government consumption share	330	0.291	0.061	0.171	0.479
SOE employment share	330	0.125	0.061	0.053	0.428

Notes: Data cover 30 provinces during the period 2000–2010, sourcing from various statistical yearbooks explained in Table 4.4.

4.4.2.1 Main variables

Land conflicts. The dependent variable, denoted “land conflicts”, is measured by the illegal land use area, as explained in Section 4.2. Although illegal land use is the best available proxy for land conflicts, it may suffer from measurement errors because illegal land users have incentives to conceal their illegal behavior. If such concealment is proportional to the illegal land use reported in the yearbooks, this will make the measure (and thus the estimated coefficients of the explanatory variables) systematically downward biased. The true values would then be larger than our estimates. To limit the impact of potential measurement errors, we include indicators of local institutional quality as control variables in the model.

Foreign direct investment. FDI is measured by annual FDI inflows per capita. We normalize FDI by the population size of each province because per capita FDI reflects regional capital intensity, which is more relevant to the dependent variable than the absolute value of FDI. We use current FDI instead of the accumulated stock of FDI because investments that took place many years ago are unlikely to affect current land conflicts. Figure 4.2 shows the spatial distribution of FDI during the 1999–2010 period. As expected, the highest concentration of FDI during that period can be found in provinces in eastern and southern China, including Beijing and Guangdong. High levels can also be found in other eastern provinces, such as Shandong and Jiangsu, and in provinces in the northeast, including Inner Mongolia.

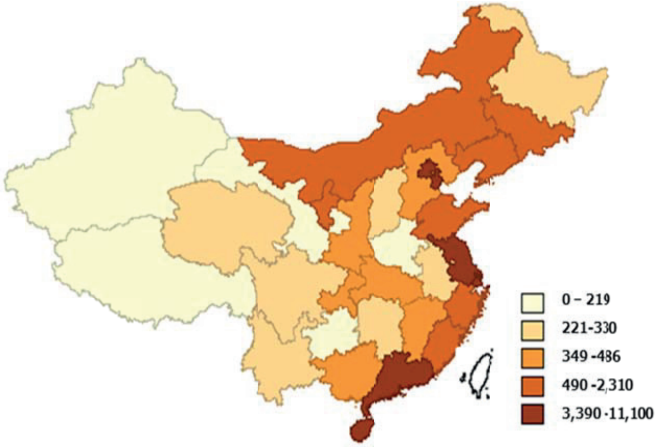


Figure 4.2 Spatial Distribution of FDI Per Capita, 1999-2010 (annual averages, 2005 US$)

Notes: FDI per capita stands for foreign direct investment inflows, weighted by provincial population. Its values, in constant 2005 U.S. dollars, are calculated using statistics from China Statistical Yearbook for Regional Economy.

Fiscal decentralization. Multiple measures of fiscal decentralization have been used in the literature. In both international and sub-national studies, the effects of fiscal decentralization on economic growth are found to be sensitive to the measures used (Martinez-Vazquez and McNab, 2003; Thornton, 2007; Akai and Sakata, 2002; Qiao et al., 2008). In this paper, we use two conventional measures. One (hereafter referred to as FD1) is related to fiscal autonomy or self-reliance and is defined as the ratio of local fiscal revenue to government expenditure. The other (hereafter referred to as FD2) is an indicator of the relative size of local fiscal revenues compared with central revenues and is defined as local fiscal revenues per capita divided by the sum of local fiscal revenues per capita and central fiscal revenues per capita. We use both indicators to check the robustness of our results. Figures 4.3a and 4.3b show the spatial distribution of these two indicators. They suggest that fiscal decentralization was highest in the coastal provinces but that the choice of indicator is important. Using the second indicator, Xinjiang and Inner Mongolia are found to have high degrees of fiscal decentralization, whereas the first indicator suggests that they both have relatively low levels of fiscal decentralization. Another measure that is also widely used in the literature focuses on the expenditure side of decentralization. This measure is calculated using fiscal expenditure data rather than fiscal revenues. We do not use this indicator here because local fiscal expenditure is partly financed by central fiscal transfers, and local governments may lack sufficient autonomy in its allocation.

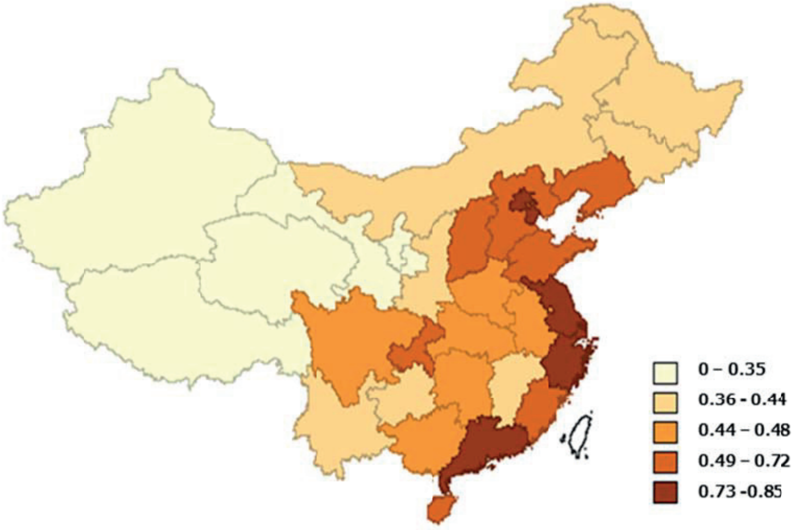


Figure 4.3a Spatial Variations in Fiscal Decentralization 1, 1999-2010 (annual averages)

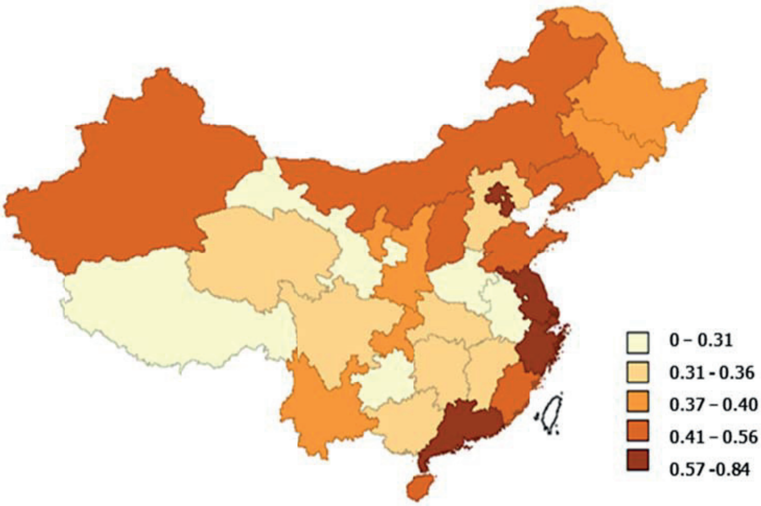


Figure 4.3b Spatial Variations in Fiscal Decentralization 2, 1999-2010 (annual averages)

Notes: Fiscal decentralization 1 is defined as the ratio of local fiscal revenue to government expenditure, while fiscal decentralization 2 is defined as local fiscal revenues per capita divided by the sum of local fiscal revenues per capita and central fiscal revenues per capita. They are both calculated using statistics obtained from China Statistical Yearbook.

Land finance. Revenues obtained from acquiring rural land and leasing out urban land are not a formal fiscal source like other budgetary items. We define land finance as the ratio of land leasing revenues divided by local budgetary fiscal revenues. In this sense, it can be interpreted as the local government's dependence on land leasing revenues relative to formal fiscal budgetary revenues. Unfortunately, information on the cost of acquiring and preparing land is not available. Hence, the land finance variable in our model reflects the revenues obtained for urban land leasing but not the profit made by local governments on such transactions.

4.4.2.2 Control variables

Regional economic development. We control for the effects of regional economic development on (the growth rate of) land conflicts by using the urbanization rate, the economic growth rate, and (the log of) per capita GDP as control variables in the regressions.

Institutional quality. Institutional quality is intended to capture local efforts in implementing land use policies and laws. One proxy we use is the share of government consumption in total consumption. It reflects the degree of government intervention in the economy and is negatively correlated with economic freedom (Gwartney et al., 2013). The other measure we use is the ratio of State-owned enterprise (SOE) employment to total wage employment.[32]

4.5 Results and sensitivity checks

4.5.1 Baseline results

The regression results of model (1) are shown in Table 4.6. Columns (1) and (3) show the results without the interaction term for the two fiscal decentralization measures (FD1 and FD2, respectively), while columns (2) and (4) present the results when the interaction term is included. The main finding is that the estimated coefficients of FDI and fiscal decentralization do not differ significantly from zero when the interaction term is not included; however, when the interaction term is added to the model, both the growth rate of FDI and its interaction with fiscal decentralization have highly significant coefficients. This finding is independent of the choice of fiscal decentralization measure. It provides support for our main hypothesis (see Section 4.3.4) that the interaction between FDI and fiscal decentralization contributes positively to land conflicts.

[32] Wage employment refers to employment based on a work unit, i.e.,"danwei jiuye", in the NBS statistics.

Table 4.6: Regression results for land disputes, OLS fixed effects

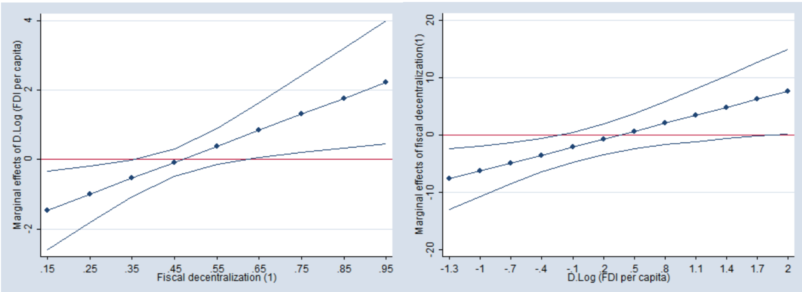
	(1)	(2)	(3)	(4)
	FD1	FD1	FD2	FD2
Δln (FDI per capita)	-0.289	-2.158**	-0.281	-3.323**
	(0.238)	(0.836)	(0.241)	(1.506)
Fiscal decentralization	-1.282	-1.696	-2.392	-2.801
	(1.281)	(1.294)	(2.377)	(2.708)
Δln (FDI per capita) *		4.619**		7.597**
Fiscal decentralization		(1.789)		(3.635)
Land finance	0.107	-0.009	0.135	0.133
	(0.193)	(0.190)	(0.203)	(0.204)
GDP growth rate	7.867***	7.521***	8.766***	8.810***
	(1.887)	(2.042)	(1.945)	(2.553)
ln (GDP per capita)	-0.898	-1.019*	-0.601	-0.911
	(0.573)	(0.548)	(0.819)	(0.801)
Urbanization rate	-0.196	-0.196	-0.460	-0.342
	(1.756)	(1.689)	(1.847)	(1.894)
Government consumption ratio	0.284	0.680	0.724	1.145
	(1.127)	(1.136)	(1.172)	(1.187)
SOE employment share	-4.130	-4.047	-3.381	-2.942
	(3.331)	(3.201)	(3.200)	(3.072)
Year-fixed effects	Yes	Yes	Yes	Yes
Province-fixed effects	Yes	Yes	Yes	Yes
Number of observations	330	330	330	330
Adjusted R2	0.181	0.197	0.181	0.197

Notes: Robust standard errors clustered at provincial level in brackets; *, * *, *** denote significance levels of 10%, 5%, 1%, respectively. The dependent variable is the change in the logarithm of illegal land use area. Columns 1-2 use the variable FD1 as the measure of fiscal decentralization, while columns 3-4 use FD2. Δln (FDI per capita) * Fiscal decentralization represents the interaction term between Δln (FDI per capita) and Fiscal decentralization. Variable definitions are described in detail in Table 4.4 and the text. All specifications are estimated using OLS, controlling for both year and province fixed effects.

To visualize the interaction effects between FDI and fiscal decentralization, we derive the marginal changes in the dependent variable in response to a one-unit change in the fiscal decentralization or FDI variables. In Figures 4.4 and 4.5, panel A display the marginal effects (with 95% confidence intervals) of the growth rate of FDI at given levels of fiscal decentralization, as derived from the two models with interaction terms.[33] Both graphs show that FDI has a significant, negative effect on

[33] The marginal effects estimates are generated with Stata (version 13) command "margins" given after the regressions for which the results are presented in columns 2 and 4 in Table 4.6. As an example, Panel A of Figure 4.4 is generated in the following way. First, the command "margins" calculates the predicted change in the

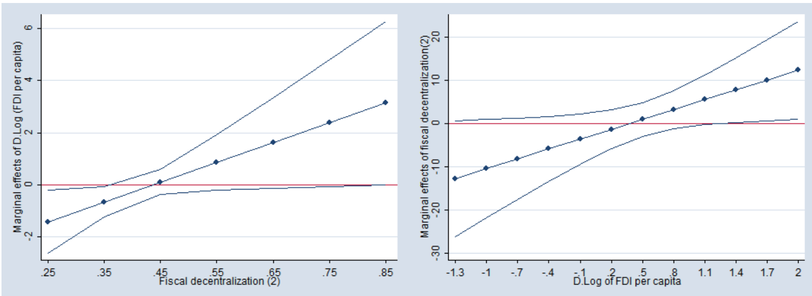
land conflicts at low levels of fiscal decentralization; at high levels of fiscal decentralization, however, the impact of FDI becomes significantly positive. Panel B in the same two figures shows the marginal effects of fiscal decentralization on land conflicts for different FDI growth rates. When FDI growth rates are low, fiscal decentralization has significantly negative effects on land conflicts; however, when FDI growth rates are high, fiscal decentralization significantly contributes to the occurrence of land conflicts.



(a) Panel A **(b)** Panel B

Figure 4.4: Marginal Effects of Δln (FDI per capita) and Fiscal Decentralization 1

Notes: The point estimates are calculated on the basis of model 2 in Table 4.6. The lower and upper bounds give the 95% confidence interval. Footnote 7 explains the detailed procedures.



(a) Panel A **(b)** Panel B

Figure 4.5: Marginal Effects of Δln (FDI per capita) and Fiscal Decentralization 2

Notes: The point estimates are calculated on the basis of model 4 in Table 4.6. The lower and upper bounds give the 95% confidence interval. Footnote 7 explains the detailed procedures.

Another important finding is that the estimated coefficient of land finance does not differ significantly from zero in all the four regressions. This result suggests that farmers are less likely to be engaged in land conflicts when their land is used as

---

dependent variable in a certain province-year when the rate of FDI change increases by one unit while holding FD1 constant and using the observed values of the other explanatory variables for each province-year. Second, "margins" calculates the means and standard deviations over all province-years of the predicted marginal changes. Steps one and two are repeated for values of FD1 in the range 0.15–0.95 (i.e., the sample range of FD1). Finally, "marginsplot" plots the means and confidence intervals.

residential land or leased out to the services sector. A possible explanation, which requires further research, is that farmers receive higher compensations when their expropriated land is leased to the tertiary sector than when leased to the secondary sector. It should be noted, however, that the land finance variable in our analysis does not take the costs of acquiring and preparing land into account. It would be interesting to investigate whether our finding still holds when this measurement error is corrected (once appropriate data for doing so become available).

Regarding the control variables, the results in all four columns consistently suggest that the economic growth rate has significant, positive effects on land conflicts. The positive coefficient for economic growth differs from the findings of some international studies in which economic growth is found to be a means of avoiding conflict (e.g., Miguel et al., 2004). A possible explanation for our finding is that, in the case of China, expropriated farmers pay the cost of economic development but do not enjoy a fair share of the benefits. They observe other groups in society benefitting from rapid economic growth but hardly profit themselves.

4.5.2 Sensitivity checks

This section provides sensitivity checks with respect to two concerns. One concern is that land conflicts could be related to past or even future FDI inflows. On the one hand, there may be a time lag between land expropriation for FDI and the land conflicts that arise from it. This may be caused, for example, by arrears in compensation payments. On the other hand, land may be expropriated in anticipation of foreign investments coming to the region. Hence, land conflicts may precede the actual FDI. A straightforward way to address this concern is to introduce lags and leads of FDI into the regression models. However, including leads of FDI will introduce further unobserved noise into the error term. It will cause violation of the i.i.d. assumption of the error term. Instead, we opt to average all variables in the model over either two or three years to smooth out cyclical fluctuations. Using smoothed data will allow us to examine whether the medium-term relationship between FDI and land conflicts, as represented by the 3-year smoothed averages, differs from the short-term relationship. We discuss this approach and present its results in Section 4.5.2.1.

The other concern is that the baseline regressions do not consider the potential endogeneity of FDI. There are two potential sources of endogeneity. One is omitted variables that are correlated both with the explanatory variables and with the dependent variable in the model. In Section 4.4.1 we argue that, by using fixed-province and fixed-year effects in our panel data set and taking log growth rates of the main variables of interest, this source has largely been eliminated. Another potential source of endogeneity is reverse causality. We address this issue in Section 4.5.2.2 and in the Appendix A.

4.5.2.1 Results for moving averages

Table 4.7 presents the regression results of model (1) with smoothed data instead of annual data. Columns (1) – (4) display the results with 2-year moving averaged data; columns (5)–(8) report the results for 3-year moving averaged data. Smoothed data mitigate the effects of cyclical fluctuations in FDI, among other predictors. Interestingly, the results from the 2-year moving averaged data are broadly consistent with the baseline regression results. The estimated coefficients of the interaction term are somewhat larger in size than those estimated with the annual data. However, when we use 3-year averaged data, we find that the estimated coefficients of the FDI growth rate and the interaction term are not statistically significant. These results suggest that the positive impact of the interaction between FDI and fiscal decentralization on land conflicts is mainly a short-run relationship. The relationship is not significant when we adopt a medium-term (3-year) horizon, thereby suggesting the presence of countervailing feedback effects after a few years. Possible explanations are that local land disputes strengthen the resistance against land requisitions and thus make it more difficult to use land as an instrument for attracting FDI or that foreign investors are scared away by land conflicts.

Further note that, again, no significant impact is found for the land finance variable. The estimated coefficient of the GDP growth rate, however, is significant in 6 out of the 8 models shown in Table 4.7. Its magnitude is somewhat smaller in the models with smoothed averages than in that using annual data. However, even using these estimates, we find that economic growth has a sizable, positive impact on land disputes.[34]

Table 4.7: Regression results for land disputes: OLS fixed effects using moving averages

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Panel A: 2-year average data				Panel B: 3-year average data			
	FD1	FD1	FD2	FD2	FD1	FD1	FD2	FD2
Δln (FDI per capita)	-0.124	-1.663**	-0.110	-3.374*	-0.164	-1.090	-0.142	-2.871
	(0.182)	(0.742)	(0.178)	(1.745)	(0.163)	(0.766)	(0.162)	(1.831)
Fiscal decentralization	-1.134	-1.418	-0.916	-1.394	-0.928	-1.043	-1.390	-1.767
	(1.115)	(1.102)	(2.126)	(2.307)	(0.956)	(0.916)	(1.971)	(1.986)
Δln(FDI per capita)* Fiscal decentralization		3.774**		8.212*		2.236		6.819
		(1.579)		(4.495)		(1.672)		(4.691)
Land finance	0.335	0.216	0.360	0.372	0.391	0.297	0.419	0.448
	(0.230)	(0.225)	(0.238)	(0.244)	(0.266)	(0.270)	(0.287)	(0.306)
GDP growth rate	5.710**	5.258**	6.564***	5.985**	5.592**	5.322**	6.429**	5.819**
	(2.226)	(2.445)	(2.098)	(2.386)	(2.383)	(2.557)	(2.377)	(2.493)
ln (GDP per capita)	-0.814*	-0.889*	-0.844	-1.085*	-1.029**	-1.118**	-0.894	-1.131*
	(0.474)	(0.487)	(0.646)	(0.625)	(0.472)	(0.501)	(0.627)	(0.623)
Urbanization rate	0.162	0.224	0.077	0.285	0.674	0.723	0.537	0.817
	(1.769)	(1.690)	(1.915)	(1.869)	(1.880)	(1.823)	(1.891)	(1.792)
Government consumption ratio	0.226	0.628	0.516	0.948	0.779	1.095	1.098	1.520
	(1.101)	(1.105)	(1.180)	(1.180)	(1.107)	(1.147)	(1.161)	(1.203)
SOE employment share	-3.101	-2.760	-2.428	-1.749	-2.345	-1.818	-1.746	-0.643
	(3.449)	(3.422)	(3.205)	(3.207)	(3.241)	(3.452)	(3.021)	(3.349)
Year-fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Province-fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of observations	300	300	300	300	270	270	270	270
Adjusted R2	0.248	0.261	0.245	0.265	0.262	0.269	0.262	0.277

Notes: Robust standard errors clustered at provincial level in brackets; *, * *, *** denote significance levels of 10%, 5%, 1%, respectively. The dependent variable is the change in the logarithm of illegal land use area. Panels A and B employ 2-year and 3-year moving averaged data, respectively, to smooth out cyclical fluctuations. Columns 1, 2, 5, and 6 use the variable FD1 as the measure of fiscal decentralization, while columns 3, 4, 7 and 8 use FD2. Δln (FDI per capita) * Fiscal decentralization represents the interaction term between Δln (FDI per capita) and Fiscal decentralization. Variable definitions are described in detail in Table 4 and the text. All specifications are estimated using OLS, controlling for both year and province fixed effects.

4.5.2.2 Endogeneity

The intensity of local land conflicts may play a role in FDI location choice. If such feedback effects are present, the estimates obtained thus far may be biased. A simple, although not conclusive, way to test for reverse causality is to regress (the rate of

[34] Assuming that the estimate coefficient of the GDP growth rate equals 5.0 (based on the estimates in Table 4.6), a one-standard-deviation change in the GDP growth rate (2.4%, see Table 4.4) will increase the growth rate of illegal land use by 2.4%*5=12%. This explains approximately 14% of the standard deviation of illegal land use growth (which equals 85.8%).

change in) FDI on one-year and two-year lagged land conflicts and a number of control variables. The results are shown in Table 4.8. The estimated coefficient of land conflicts is not statistically significant in any of the four models. We also do not find significant effects for almost all control variables. When we vary the specification of the control variables in the model, we obtain similar results.[35]

Table 4.8: Regression results using FDI as the dependent variable, OLS fixed effects

	(1)	(2)	(3)	(4)
	FD1	FD2	FD1	FD2
1-year lag of Δln (illegal land use area)	0.006	0.006	-0.003	-0.002
	(0.016)	(0.016)	(0.021)	(0.021)
2-year lag of Δln (illegal land use area)	-	-	-0.021	-0.023
	-	-	(0.019)	(0.021)
Fiscal decentralization	-0.607	-0.606	-0.759	-1.221
	(0.658)	(1.594)	(0.770)	(2.101)
Land leasing share	-0.080	-0.068	-0.068	-0.052
	(0.077)	(0.075)	(0.084)	(0.084)
Growth rate of real GDP	1.361	1.764	1.206	1.804
	(1.119)	(1.182)	(1.038)	(1.230)
ln(real GDP per capita)	-0.286*	-0.262	-0.374*	-0.250
	(0.157)	(0.338)	(0.203)	(0.410)
Urbanization rate	0.605	0.555	1.280	1.190
	(0.746)	(0.734)	(1.041)	(1.034)
Government consumption share	0.489	0.642	0.514	0.781
	(0.353)	(0.460)	(0.433)	(0.568)
SOE employment share	0.806	1.179	1.468	2.245
	(0.688)	(0.755)	(1.262)	(1.524)
Year-fixed effects	Yes	Yes	Yes	Yes
Province-fixed effects	Yes	Yes	Yes	Yes
Observations	300	300	270	270
R-squared	0.084	0.080	0.010	0.098

Notes: Robust standard errors clustered at provincial level in brackets; *, * *, *** denote significance levels of 10%, 5%, 1%, respectively. The dependent variable is the change in the logarithm of FDI per capita. Columns 1 and 3 use the variable FD1 as the measure of fiscal decentralization, while columns 2 and 4 use FD2. Columns 1 and 2 include the one-year lagged change rate of illegal land use area as an explanatory variable, while columns 3 and 4 further add its two-year lag. Other variable definitions are described in detail in Table 4.4 and the text. All specifications are estimated using OLS, controlling for both year and province fixed effects.

A further step we take to formally address potential endogeneity issues in our model is to apply Generalized Method of Moments (GMM) estimators. We use both the "system" GMM and the "difference" GMM method. Details of their specifications and estimation results can be found in the Appendix. We find that the estimated coefficients for the FDI * fiscal decentralization interaction term are positive and significantly different from zero in four out of the eight estimated models. The

[35] These results are available upon request from the first author.

combined impact of FDI and fiscal decentralization on land conflicts is thus only partially supported by the GMM evidence. This is not at odds with expectations, as the assumptions made with respect to the instruments are likely to affect the results and the finding that the two GMM estimation procedures can produce strikingly different results (e.g., Bobba and Coviello, 2007).

4.6 Conclusion

Land conflicts have attracted increasing policy attention in contemporary China. However, systematic academic studies of the underlying causes of land conflicts remain sparse. In this paper, we argue that FDI inflows and fiscal decentralization play important roles in generating land conflicts. FDI can stimulate better governance and create more employment and, thereby, attenuate land conflicts. However, when the degree of fiscal decentralization is high, local governments tend to become involved in a race to the bottom with other local governments in their efforts to attract FDI by offering low land prices and other amenities. Fiscal decentralization further stimulates local governments to raise more "extra-budgetary" revenues by offering low land prices to farmers and leasing the land for high prices to the tertiary sector in urban areas. Both channels lead to more land conflicts.

The empirical results generally support our argument that the joint effect of FDI and fiscal decentralization is to amplify the number of land conflicts. Regional FDI inflow has a negative effect on land conflicts when the level of fiscal decentralization is low. However, the effect becomes positive and increases as the degree of fiscal decentralization increases. Moreover, the economic growth rate also tends to have significantly positive effects on land conflict. However, we do not find evidence that the dependence on land leasing revenue has significant, positive effects on land conflicts.

The results imply that FDI in particular and globalization in general are double-edged swords for Chinese economy. FDI may positively affect overall economic growth but may also have negative distributional effects when there are groups whose political and property (farmers' land in our case) rights are not well protected. The worsening of their (relative) welfare carries seeds of protest, revolution, and military conflict (Acemoglu and Yared, 2010; Martin et al., 2008). Such negative distributional effects are also reported in recent research in other developing economies in East Asia, Latin America, and Africa (Goldberg and Pavcnik, 2007). One interesting study by Bezemer and Jong-A-Pin (2013) shows that globalization increases ethnic violence in economies dominated by minority groups.

Our findings suggest that social tension over land in China is also among the side effects of globalization and governmental dominance over the local economy. To manage the gains of land value booms in China, fiscal arrangements may be further improved by re-classifying local governments' revenues sources and expenditure responsibilities to limit their incentives to engage on land finance. The rules of land requisition may need to be strengthened to effectively protect the land rights of farmers (and their collective organizations) and to allow farmers to benefit from the incremental value created by assembling and developing the land.

The analysis of this paper is limited by the information available on land conflicts. While the empirical examination is based on provincial-level data, it would be preferable to have lower-level regional data to better control for regional heterogeneity. Our proxy for land conflicts, i.e., illegal land use, only represents the formal land disputes addressed via legal processes. It is thus not representative of other types of land disputes such as land-triggered mass incidents. In addition, we only slightly touched upon the potential role of windfall gains in land values in generating land conflicts. The nature of our data does not allow us to conduct a systematic examination. Future studies may find micro-level survey data particularly useful to improve our understanding on how land values gains reshape local society.

Appendix A: Dynamic patterns, endogeneity, and GMM estimations

We apply standard Generalized Method of Moments (GMM) estimators to account for both the potential endogeneity of the independent variables and dynamic patterns in illegal land use. To examine potential convergence effects, we add one-year-lagged illegal land use area (log) on the right-hand side of equation (1) and estimate it via GMM. Arellano and Bond (1991) first proposed the GMM estimator for dynamic panel data models. It takes the first difference of the data and then uses lagged values of the endogenous variables as instruments (the so-called "difference" GMM estimator). However, as noted by Arellano and Bover (1995), lagged levels are weak instruments for first differences when the explanatory variables are persistent over time. Blundell and Bond (1998) proposed the "system" GMM estimator that includes additional moment conditions in level equations. Though the system GMM estimator is more efficient, it is more subject to the problem of having too many instruments than is the difference GMM estimator (Hayakawa, 2007; Asiedu and Lien, 2011). As shown in the literature, the two estimation procedures may produce strikingly different results (e.g., Bobba and Coviello, 2007). We apply both the difference and system GMM estimators.

We discuss the detailed specifications of our GMM estimations, as results are often sensitive to these choices (Bezemer and Jong-A-Pin, 2013). Particular attention is devoted to the issue of having "too many instruments" as numerous instruments can "overfit" endogenous variables, leading to biased estimators. First, we treat all the independent variables, except the year dummies, as endogenous in all the regressions. Second, we limit the length of the time lag depth to 2–3 years for each explanatory variable.[36] Thus, both the difference and system estimators use the first difference of all the exogenous variables (year dummies) as standard instruments and the 2nd–3rd lags of the endogenous variables to generate the GMM-type instruments. The system estimations additionally include lagged differences of the endogenous variables as instruments for the level equation. Third, we collapse instruments in all estimations to further reduce the size of the instrument sets. The

[36] If the lag depth is further limited to 2, the Sargan test will fail because the number of (collapsed) instruments equals the number of parameters to be estimated.

choice of lags and the use of collapsed instruments are intended to limit the number of instruments to a number that is smaller than that of the individual units as suggested by Roodman (2009). In so doing, we are able to have 29 instruments in the difference GMM estimations but 40 instruments for the system GMM estimations (see Table 4.A1). It is possible to further reduce the size of the instrument sets, for example by dropping the year fixed effects from the regressions (e.g., Asiedu and Lien, 2011). Tables 4.A1 and 4.A2 present the results when including and excluding the year dummies in the model, respectively.

The results with year dummies presented in Table 4.A1 show that the Sargan test and the Arellano–Bond test do not reject the validity of the results. The p-values for the Sargan test reported for all regressions are larger than 0.378, which suggests that there is no evidence to reject the validity of the instrumental variables.[37] We also report the Arellano–Bond test results for serial correlation in the first-differenced errors. The test statistics (AR2 test) do not provide evidence that the model is misspecified (the p-values range between 0.096 and 0.289). When we focus on the interaction term between FDI and fiscal decentralization, we find that the coefficients are not consistently significant. The coefficients reported in columns 2, 4, 5 and 6 are statistically significant, and their signs are consistent with those of the baseline regressions. The impact of FDI and fiscal decentralization is thus only partially supported by the GMM evidence. The sensitivity of the main results to the use of GMM approaches may be because our panel data set has a relatively small number of cross-sectional observations N and a relatively large number of annual observations T. In this case, the instrument set tends to be large (see also Zheng et al., 2013). In Table 4.A2, we apply the method of excluding year dummies from the model to reduce the size of the instrument sets. The positive interaction effects are statistically significant in 6 out 8 regressions.

[37] Sargan's over-identification restriction test is less vulnerable to the number of instruments. The Sargan test is found to be more conservative than the Hansen test. We thus choose to report the Sargan test statistics.

Table 4.A1: Regression results for land disputes: Difference and System GMM

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	GMM-DIF	GMM-SYS	GMM-DIF	GMM-SYS	GMM-DIF	GMM-SYS	GMM-DIF	GMM-SYS
	FD1	FD1	FD2	FD2	FD1	FD1	FD2	FD2
	Panel A: One-step Estimator (Collapsed IV matrix)				Panel B: Two-step Estimator (Collapsed IV matrix)			
log (illegal land use area$_{t-1}$)	-0.944***	-0.758***	-0.858**	-0.716***	-0.982***	-0.676***	-0.795***	-0.583***
	(0.219)	(0.170)	(0.355)	(0.182)	(0.123)	(0.198)	(0.207)	(0.156)
Δlog (FDI per capita)	-5.857	-4.885**	-6.288	-5.301**	-4.804**	-4.661**	-8.450**	-2.506
	(3.584)	(2.202)	(5.501)	(2.201)	(2.142)	(1.832)	(4.281)	(2.313)
Fiscal decentralization	1.272	0.358	-33.718	-2.920	1.184	0.219	-36.088**	-11.070*
	(4.276)	(1.988)	(22.076)	(4.177)	(3.365)	(2.955)	(16.039)	(6.531)
Δlog (FDI per capita) * Fiscal decentralization	11.735	9.481*	10.871	10.137*	8.128*	8.129**	14.641	2.992
	(7.914)	(5.079)	(12.497)	(5.205)	(4.833)	(4.066)	(9.881)	(5.316)
Land finance	0.720	0.132	0.223	-0.066	0.523	0.417	-0.343	-0.022
	(1.161)	(0.481)	(1.695)	(0.432)	(0.782)	(0.399)	(1.325)	(0.471)
GDP growth rate	16.472	10.107	18.323	9.385	10.014	19.392*	12.929	7.389
	(11.804)	(7.122)	(15.314)	(6.261)	(8.371)	(11.726)	(10.668)	(5.433)
log (GDP per capita)	-0.417	-0.656	9.722	0.817	-0.925	-1.082	9.128**	3.759*
	(1.943)	(0.842)	(6.220)	(1.410)	(0.946)	(1.327)	(4.347)	(2.116)
Urbanization rate	0.922	0.788	-13.921	-0.400	-0.247	2.317	-6.945	-3.457*
	(7.781)	(2.257)	(10.248)	(2.083)	(5.549)	(3.782)	(7.768)	(2.062)
Government consumption ratio	-1.174	2.118	-3.185	1.139	1.625	2.051	1.473	0.763
	(5.383)	(3.687)	(8.479)	(2.440)	(3.368)	(2.886)	(6.023)	(1.998)
SOE employment share	-6.463	-2.390	-11.188	-3.347	-0.493	-2.308	-1.430	2.566
	(9.880)	(4.419)	(12.494)	(3.485)	(6.633)	(5.746)	(8.957)	(3.436)
Year-fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Lags used	2nd-3rd	2nd-3rd	2nd-3rd	2nd-3rd	2nd-3rd	2nd-3rd	2nd-3rd	2nd-3rd
AR(2) test p value	0.242	0.264	0.174	0.213	0.119	0.268	0.096	0.289
Sargan test p value	0.587	0.425	0.947	0.378	0.587	0.425	0.947	0.378
Number of instruments	29	40	29	40	29	40	29	40
Number of provinces	30	30	30	30	30	30	30	30
Number of observations	270	300	270	300	270	300	270	300

Notes: Robust standard errors in brackets; *, * *, *** denote significance levels of 10%, 5%, 1%, respectively. The dependent variable is the change in the logarithm of illegal land use area, whereas log (illegal land use area$t-1$) denotes its one year lag. GMM-DIF and GMM-SYS correspond difference GMM and system GMM estimation methods, respectively. Specifications in Panel A are estimated with one-step GMM estimator, while specifications in Panel B are estimated with two-step estimator. Columns 1, 2, 5 and 6 use the variable FD1 as the measure of fiscal decentralization, while columns 3, 4, 7 and 8 use FD2. All the left hand side variables, except year dummies, are treated as endogenous; their 2nd-3rd lags are used to generate the GMM-type (collapsed) instruments.

Table 4.A.2: Regression results for land disputes: Difference and System GMM, without year fixed effects

	(1) GMM-DIF	(2) GMM-SYS	(3) GMM-DIF	(4) GMM-SYS	(5) GMM-DIF	(6) GMM-SYS	(7) GMM-DIF	(8) GMM-SYS
	Panel A: One-step Estimator (Collapsed IV matrix)				Panel B: Two-step Estimator (Collapsed IV matrix)			
log (illegal land use area$_{t-1}$)	-1.078***	-0.864***	-0.961***	-0.786***	-1.034***	-0.854***	-1.007***	-0.731***
	(0.192)	(0.170)	(0.189)	(0.175)	(0.122)	(0.043)	(0.134)	(0.067)
Δlog (FDI per capita)	-5.971*	-4.412**	-10.533	-5.901***	-5.777**	-3.884***	-10.723***	-5.450***
	(3.462)	(1.771)	(6.508)	(2.158)	(2.339)	(0.667)	(4.076)	(1.261)
Fiscal decentralization	-0.608	-0.162	4.321	-0.158	-2.052	-0.754	1.853	-0.663
	(3.729)	(1.249)	(10.031)	(1.780)	(3.096)	(0.665)	(6.643)	(1.006)
Δlog (FDI per capita) * Fiscal decentralization	12.253	9.313**	22.766	12.196**	12.002**	8.095***	23.026**	10.584***
	(7.485)	(3.998)	(14.704)	(4.998)	(5.153)	(1.734)	(9.183)	(2.739)
Land finance	0.952	0.150	1.508*	0.175	1.480***	0.527**	1.263**	0.314**
	(0.623)	(0.429)	(0.869)	(0.343)	(0.428)	(0.235)	(0.550)	(0.154)
GDP growth rate	13.147**	12.256**	11.697	13.369***	11.768***	12.512***	10.894***	13.432***
	(5.903)	(5.124)	(7.406)	(4.173)	(3.608)	(2.095)	(4.163)	(1.570)
log (GDP per capita)	-0.477	-0.259	-1.039	-0.356	-0.661**	-0.313**	-0.841	-0.329***
	(0.413)	(0.238)	(0.900)	(0.219)	(0.282)	(0.133)	(0.537)	(0.119)
Urbanization rate	-4.140	1.036	-3.825	1.083	-4.647	2.008**	-3.060	0.318
	(7.540)	(2.124)	(5.382)	(1.765)	(4.592)	(0.909)	(3.813)	(0.949)
Government consumption ratio	-4.171	1.316	-8.600	1.423	-4.312	0.937	-5.645	2.353*
	(4.549)	(3.047)	(5.770)	(2.668)	(3.035)	(1.439)	(4.102)	(1.265)
SOE employment share	-12.193**	-4.400	-19.348**	-4.937	-12.652***	-1.522	-15.304***	-1.712
	(4.996)	(3.696)	(8.649)	(3.243)	(3.898)	(1.625)	(4.693)	(1.291)
Year-fixed Effect	No	No	No	No	No	No	No	No
Lags used	2nd-3rd	2nd-3rd	2nd-3rd	2nd-3rd	2nd-3rd	2nd-3rd	2nd-3rd	2nd-3rd
AR(2) test p value	0.096	0.111	0.067	0.096	0.062	0.056	0.061	0.051
Sargan test p value	0.556	0.051	0.712	0.031	0.556	0.051	0.712	0.031
Number of Instruments	20	31	20	31	29	40	29	40
Number of Provinces	30	30	30	30	30	30	30	30
Number of Observations	270	300	270	300	270	300	270	300

Notes: Robust standard errors in brackets; *, * *, *** denote significance levels of 10%, 5%, 1%, respectively. The dependent variable is the change in the logarithm of illegal land use area, whereas log (illegal land use area$_{t-1}$) denotes its one year lag. GMM-DIF and GMM-SYS correspond difference GMM and system GMM estimation methods, respectively. Specifications in Panel A are estimated with one-step GMM estimator, while specifications in Panel B are estimated with two-step estimator. Columns 1, 2, 5 and 6 use the variable FD1 as the measure of fiscal decentralization, while columns 3, 4, 7 and 8 use FD2. All left-hand-side variables are treated as endogenous; their 2nd–3rd lags are used to generate the GMM-type (collapsed) instruments

CHAPTER 5

The Gender Wage Gap in China's Rural-Urban Migrants' Labor Market: New Evidence

Abstract

Large gender wage gaps have been found among rural-urban migrants in China in a number of empirical studies. Data used in these studies are mainly collected among long-term migrants residing in urban communities and largely exclude migrants living at their workplaces. In this paper, using a newly available data set, we aim to compare the magnitude of gender wage differentials and its sources between these two sub-groups of rural-urban migrants. The gender wage gaps that we find, 16-18%, are smaller than those documented in previous studies and do not differ much between the two sub-groups. The sources of gender wage differentials, in contrast, exhibit important differences between the two groups: gender differences in industry sorting are relatively important for the gender wage gaps of migrants living at their workplaces, whereas differences in education and experience are of particular importance for explaining the gender wage gaps of those living in urban communities. Gender differences in observed characteristics, however, explain only a small part of observed gender wage gaps among migrants, especially for the sub-group that lives in urban communities.

Keywords: Rural-urban migrants, Gender wage gap, Decomposition analysis, China

JEL classification: J31, J71

This chapter is about to be submitted to China Economic Review, as Wu, Y. , J. Pieters, N. Heerink. 'The Gender Wage Gap in China's Rural-Urban Migrants' Labor Market: New Evidence.'

5.1 Introduction

The Chinese urban labor market, largely due to the Household Registration System (the so-called *Hukou* system), has been characterized by segmentation between rural migrants and urban residents (see Meng, 2012 for a survey). Many studies document that rural-urban migrants (hereinafter referred to as migrants) in China earn much less than urban residents; according to Zhu (2015), the (raw) wage differential between urban workers and migrants amounted to 48% in 2002 and 58% in 2007. Female migrants in China's urban labor market may even face a "double disadvantage" due to male-female wage gaps. Available literature, however, has paid relatively little attention to gender wage gaps among China's migrants.

Narrowing the gender wage gap among migrants can make an important contribution to socio-economic development in China. Higher female migrants' wages will contribute to narrowing the still rather large rural-urban income gap in China. Perhaps even more importantly, raising incomes of female migrants can be instrumental in improving the education and health of migrant children and have a range of other desirable returns beyond direct economic benefits. In particular, benefits could materialize through reduced separation of families by promoting migration of entire families instead of single (male) persons, and by increasing women's relative decision making power in migrant households (e.g. Chang et al. 2011; Majlesi 2016).

The few studies that examined gender wage gaps among Chinese migrants found sizeable differences. Magnani and Zhu (2012), for instance, report that male migrants earn 30.2% more per hour than female migrants. As regards the factors contributing to this gender wage gap, it has been found that 67 – 88 percent (depending on the studies) of the wage gap cannot be explained by gender differences in observable characteristics (Magnani and Zhu, 2012; Qin et al. 2016). Among the observable characteristics, human capital endowment and industry of employment are identified as contributing most to the gender gap. Lower relative returns to potential experience (age) of women as compared to men, especially for the higher income brackets, are another major factor contributing to the gender wage gap.

The data sets and decomposition methodologies used in this field face a few important limitations in terms of coverage and methodology. Migrants living at their workplace, usually working in construction and manufacturing industries, are underrepresented in previous studies (Demurger et al., 2009; Magnani and Zhu, 2012). As regards methodology, existing studies overwhelmingly rely on the Blinder-Oaxaca (B-O) decomposition method to examine the contributions to gender wage gaps of wage differentials caused by observable characteristics and by gender-specific wage responses to these characteristics. More recent studies, e.g., Elder et al. (2010), have offered new guidelines — but received little attention — in dealing with the sensitivity of B-O decompositions. The proposed methodology in those studies can help investigate the robustness of the findings in previous studies.

This study therefore aims to make two contributions to the literature. First, we investigate the gender wage gap and its contributing factors using new and more representative migrant data. Our data sets, drawn from the Rural-Urban Migration in China (RUMiC) project, consist of both migrants living at their workplace and those living in urban communities. Second, to examine the factors that contribute to the gender wage gap among migrants and robustness of the results, we apply both the conventional B-O decompositions and a recently developed decomposition method by Gelbach (2016).

Our main findings can be summarized as follows. First, the gender wage gaps among migrants are smaller in the RUMiC data as compared to what has been found in other studies and do not differ much between migrants living at their working places and migrants residing in urban communities. Consistent with previous findings, the so-called unexplained part accounts for most of the gender wage gap. Second, the results from different decomposition methods are comparable and robust. In particular, lower levels of education and off-farm experience, employment in lower-paying industries, and lower returns to age for females are among the most important factors contributing to the gender wage gap. Third, we also find interesting differences in gender wage determinations between the sample of migrants living at their workplaces and those living in urban communities. In particular, gender differences in industrial sorting only play a significant role for the subsample living at their workplaces, whereas differences in education and off-farm experience are more important for those living in urban communities. The total contributions from endowments are much higher in the subsample living at workplaces than in the subsample living in urban communities.

The remainder of the paper is organized as follows. The next section provides a review of the relevant literature. Section 5.3 introduces the RUMiC data set, calculates the gender wage gaps for the whole sample and several sub-samples, and presents summary statistics on the major variables used in the decomposition analyses. Section 5.4 discusses the two wage gap decomposition methods that are employed for examining the factors contributing to the observed gender wage gaps. Section 5.5 presents and discusses the empirical results. Finally, the main findings are summarized and discussed in Section 5.6.

5.2 Literature review

A well-documented finding across the globe is that men earn significantly more than women. Numerous studies have been devoted to detecting the causes of these gender earning differentials. Using the classic Blinder-Oaxaca decomposition method, parts of the gap are attributed to observable human capital characteristics, such as education and experience, while the remaining unexplained wage differences are treated as a measure of labor market discrimination against women. This line of studies typically finds that about half of the gender earnings gaps can be attributed to differences in human capital characteristics — the so-called endowment or composition effect (Grove et al., 2011). The other half is usually referred to as the proportion caused by gender discrimination, the so-called coefficient effect. The relative contribution of each of these two components, including detailed breakdowns of each component, has been the main focus of the field.

In the context of China, economists have paid relatively little attention to the gender wage gap among one major group of laborers, i.e. rural-urban migrants.

Magnani and Zhu (2012), a notable exception, find a sizable gender wage differential: male migrants earn 30% higher hourly wages than females. Their estimate is for the year 2002 and uses a sample of migrants which they call "long term migrants who can settle down in urban China". Qin et al. (2016) report that male migrants earn 26% higher hourly wages than their female counterparts, based on data drawn from the 2010 National Migrant Dynamic Monitoring Survey (NMDMS). These findings may be compared with gender earning gaps for all urban households, which ranged from 19 to 27 percent during the period 2000-2004 (Zhang et al., 2008). It means that the raw gender wage gap in China's migrants labor market is comparable to the gender wage gap among urban households in China, and also comparable to the gender gap of approximately 20-30% estimated for many developed countries (Fortin, 2005).

What is striking, however, is the fact that most of gender wage differentials among migrants are not due to the observed population characteristics. The proportions which cannot be attributed to differences in productive traits are 67% in Magnani and Zhu (2012) and 88% in Qin et al. (2016). Hence, estimated coefficient effects are larger than typical estimates of around 50% found in studies for other countries.

A major concern with the aforementioned literature is related to the data being used. Migration surveys generally apply a residence-based sampling approach, whereby interviewers randomly select migrants in specific urban neighborhoods. This common practice, however, is less appropriate for migrant surveys in China because a large group of migrants live at their workplace. These workplaces are separated from urban residence blocks and migrants living there will not enter a sample frame constructed from a residence based sampling strategy. The data we use for this study has the advantage of being a more representative migrant sample that is not restricted to migrants living in urban communities, and can be used to examine the robustness of available findings on raw and residual wage differentials between female and male migrants.

Turning to the potential explanations for gender wage gaps, well-established contributing factors include taste-based and statistical discrimination, occupational segregation, and the glass ceiling effect, among others (see Altonji and Blank, 1999). For China's migrants labor market, Magnani and Zhu (2012) show that education and industrial sorting are among the most important characteristics explaining wage gaps. By contrast, occupation sorting is found to have little impact on the gender wage gap. An earlier study of Meng and Miller (1995), using a much smaller sample, did not find a robust contribution from occupational sorting either. By contrast, occupational segregation is usually identified as an important cause of gender wage gaps in the international literature (see e.g. Altonji and Blank, 1999).

Another important finding regarding the gender wage gap among migrants in China is the presence of a glass ceiling effect (Magnani and Zhu 2012), meaning that gender wage differentials are largest at the top tail of the migrants' wage distribution. This pattern is consistent with findings documented by Albrecht et al (2003), Nordman and Wolff (2009) and Bhorat and Goga (2013) using data from Sweden, Morocco and South Africa, respectively. Manning et al. (2008) document that wage rates of males grow faster with age than those of females, especially at the early stages of career. In other words, the interruption of female careers due to, e.g., childbearing, is one common reason for the existence of gender wage gaps in many countries.

Beaudry and Lewis (2014) find that the advance in information technology increased the return to education (lowering the need for physically demanding skills), and thus contributed to the decline in the US male-female wage gap since 1980. Juhn et al. (2013, 2014) further point out that the adoption of new technology improves the relative wage and employment of women in blue-collar tasks, but not in white-collar tasks. Therefore, the evolution of technology can potentially relax the physical disadvantage of female workers.

Most of the above studies are based on B-O decompositions (Blinder 1973; Oaxaca 1973). Decomposition results using B-O methods, however, are sensitive to the choice of the non-discriminatory wage structure — the so-called index problem (see Fortin et al. 2011 for a detailed discussion). Many studies acknowledge the arbitrariness in choosing the reference wage structure, and simply report all possible decomposition results (e.g., Grove et al., 2011). Recent advances in this field indicate that the wage gap decomposition should preferably be examined with the so-called pooling strategy, i.e. regressing wages on the gender indicator together with other determinants of wages rate. The resulting estimate for the group indicator provides a unique and suitable measure of the unexplained gap (Elder et al., 2010). Based on this approach, Gelbach (2016), provides a new method to account for the contributions of each observed characteristic.

5.3 Data

5.3.1 Data source

Our data are collected by the Rural-Urban Migration in China project, which held a large scale household survey in China's major migrant receiving cities[38] since 2008. RUMiC survey data is well-suited for studying the gender wage gap among migrants due to its rich employment-related information, including employment characteristics, human capital, family backgrounds, occupation and industry composition. We refer to Akgüç et al. (2014) for a detailed description of the RUMiC survey data sets. Data for the first two waves, RUMiC 2008 and RUMiC 2009, are currently available for scientific use[39]. Note that the 2009 data were collected when the global financial crisis hit China; to minimize the impact of negative labor market shocks on our results; we use the 2008 data set only.

We restrict the sample to individuals (both wage earners and self-employed) aged 16–60 with positive earnings. Full-time homemakers, students, disabled people, retirees, and other unemployed individuals are thus excluded. This procedure results in a working sample of 6,448 individuals without missing information for relevant variables, of whom 3,943 are males (60.7%) and 2,505 are females (39.3%).

The main advantage of RUMiC data, owing to its novel sampling strategy, is its

[38] Guangzhou, Dongguan, Shenzhen, Shanghai, Nanjing, Wuxi, Hangzhou, and Ningbo in the eastern/coastal region; Zhengzhou, Hefei, Luoyang, Bengbu, and Wuhan in the central region; and Chengdu and Chongqing in the western region.

[39] The 2008 data set contains information for the calendar year 2007, whereas the 2009 data set contains information for the year 2008.

broad coverage and reliability. Instead of using a residence-based sampling approach, , the RUMiC team applied a so-called workplace based sampling method. To do so, each city is firstly divided into 500 x 500 meter blocks. A number of blocks (equal to around 12% of the sample size for each city) are then randomly selected. Next, a workplace census is conducted in each selected block. All workers present at workplaces (including the informal ones on the streets) are interviewed with questions on the industry type, the total number of workers, and the total number of migrant workers. A (full) list of migrants for the block—the sample frame—is thus created. Last, individuals are randomly selected using the frames acquired in the census stage (see Kong 2010, for more details)[40]. As a result, the RUMiC survey data have a better coverage of the whole population of migrant workers than surveys held among migrant workers living within urban communities. This is an important improvement. Our calculations, using RUMiC 2008 data, show that 2,866 migrants[41], accounting for 44% of our whole sample, live at employer-provided places like factory dormitories, back of restaurants, or construction sites.

The RUMiC data allows us to construct a subsample of migrants who live at urban communities. Results obtained for this subsample can be compared with those of studies using data from residence-based migrant surveys (e.g., data used in Magnani and Zhu, 2010). Therefore, the RUMiC data can help us to understand how differences in sampling methods affect, e.g., estimates of gender wage gaps among migrant workers.

5.3.2 Descriptive analysis

5.3.2.1 Summary statistics

Table 5.1 presents summary statistics of major relevant variables for each gender, including their mean differences. Panel A in Table 5.1 shows the statistics using the whole sample. Panels B and C present the statistics using the sample of migrants living

---

[40] Other household members are also interviewed for these selected individual migrants.

[41] All these migrants are wage earners. Self-employed migrants have no records on their living places in the survey data.

at workplaces and those in urban communities separately.[42]

Table 5.1: Characteristics of rural-urban migrants

	Female		Male		Gender difference		
	Mean	SD	Mean	SD	Mean		t -statistic
Panel A: Whole sample							
Employment & income							
Monthly income (Yuan)	1403	1014	1735	1338	-332	***	-10.62
Working hours/week	61.9	17.7	63.7	17.7	-1.81	***	-3.99
Hourly wage (Yuan)	6.06	4.28	7.27	5.49	-1.21	***	-9.36
Log (Hourly wage) (Yuan)	1.64	0.55	1.80	0.58	-0.16	***	-11.11
Human capital endowment							
Schooling (years)	8.98	2.48	9.24	2.36	-0.26	***	-4.20
Age (years)	29.8	9.25	31.5	9.96	-1.70	***	-6.85
Off-farm experience (years)	4.25	3.69	5.02	4.64	-0.77	***	-6.99
Family background							
Marital status (married = 1)	0.60	0.49	0.60	0.49	0.00		0.39
Number of Kids: 0 - 16 years	0.24	0.51	0.23	0.51	0.01		0.97
Number of Kids: 0 – 3 years	0.06	0.24	0.06	0.25	-0.00		-0.48
Observations	2505		3943				
Panel B: Living at workplace							
Employment & income							
Monthly income (Yuan)	1214	554	1494	688	-280	***	-10.62
Working hours/week	59.6	14.3	60.4	14.6	-0.83		-3.99
Hourly wage (Yuan)	5.44	2.97	6.62	3.63	-1.18	***	-9.36
Log (Hourly wage) (Yuan)	1.57	0.48	1.76	0.52	-0.18	***	-11.11
Human capital endowment							
Schooling (years)	9.16	2.32	9.34	2.3	-0.17		-4.20
Age (years)	27.4	9.3	29.8	10.3	-2.40	***	-6.85
Off-farm experience (years)	3.40	2.95	4.32	4.36	-0.93	***	-6.99
Family background							
Marital status (married = 1)	0.46	0.50	0.48	0.50	-0.02		0.39
Number of Kids: 0 - 16 years	0.07	0.27	0.07	0.28	-0.00		0.97
Number of Kids: 0 – 3 years	0.03	0.16	0.03	0.17	-0.00		-0.48
Observations	949		1917				
Panel C: Not living at workplace							

42 Besides these two groups, there are 302 migrant wage earners in the sample who claim that they receive an accommodation subsidy. It is not clear where they live, so we do not include them in the comparative analyses of migrants living at workplaces and migrants living in urban communities. The same holds for the 1,509 migrants in the RUMiC survey who report being self-employed and also did not answer the living place question in the survey.

Employment & income							
Monthly income (Yuan)	1241	552	1576	791	-335	***	-10.26
Working hours/week	55.56	13.1	57.61	14.18	-2.05	**	-3.15
Hourly wage (Yuan)	6.00	3.44	7.32	4.25	-1.33	***	-7.18
Log (Hourly wage) (Yuan)	1.67	0.48	1.85	0.53	-0.18	***	-7.42
Human capital endowment							
Schooling (years)	9.37	2.60	9.64	2.45	-0.28	*	-2.30
Age (years)	29.3	8.7	31.1	9.4	-1.81	***	-4.20
Off-farm experience (years)	3.54	2.86	4.78	4.4	-1.24	***	-6.97
Family background							
Marital status (married = 1)	0.60	0.49	0.59	0.49	0.01		-0.60
Number of Kids: 0 - 16 years	0.24	0.48	0.23	0.48	0.00		-0.19
Number of Kids: 0 – 3 years	0.06	0.23	0.07	0.27	-0.01		-0.99
Observations	850		921				

Source: Calculated from RUMiC2008. *, * *, *** denote significance levels of 10%, 5%, 1%, respectively.

As shown in panel A, on average, female migrants earned 1,403 Yuan per month in the year 2007, while their male counterparts earned 1,735 Yuan[43]. The resulting difference, 332 Yuan (20.6% of male earnings), is statistically significant at the 1% level. Although females work significantly fewer hours (1.8) than males, working hours for both female and male are high - more than 60 hours a week. The hourly wage rate is 6.06 yuan per hour for females and 7.27 yuan per hour for males – an hourly wage gap of 16.6% (in terms of male wages)[44]. The second part of panel A shows that there exist significant gender differences in human capital endowments of migrant workers: on average, female migrants have 0.26 years less schooling, are 1.7 years younger, and have 0.77 years less off-farm experience. There are no significant differences, however, in family background as can be seen from the bottom part of panel A: on average, 60% of the migrant workers are married and migrant workers have 0.06 children aged 0 – 3 years and 0.23 – 0.24 children aged 0 - 16 years.

The subsample of migrants living at workplaces in panel B shows a few noteworthy differences from the whole sample. The (mean) gender difference in working hours and education becomes smaller and insignificant. On the other hand,

[43] Monthly income is defined as earnings received in any form from the current job.

[44] Hourly wage is calculated as: monthly income / (working hours per week *4).

the gender gaps in age and experience are relatively larger than is found for the whole sample. Turning to the subsample of migrants living in urban communities (i.e., not living at workplaces) in panel C, the gender differences in employment and human capital characteristics are consistently significant. Gaps are similar to the total sample, except for the gap in experience, which is larger than in the total sample. Moreover, variables related to the family backgrounds show no significant gender differences in both panels B and C, as was found for the whole sample.

To examine differences in wages between these two subsamples, Figure 5.1 shows the kernel density distribution of log hourly wages for migrants living in urban communities and those living at workplaces. The figure shows that the wage distribution of the group living in urban communities is a bit on the right of the wage distribution of those living at workplaces. The two-sample Kolmogorov–Smirnov test rejects the null hypothesis that the logarithmic hourly wages for the two groups are drawn from the same distribution (the corrected p-value is 0.001).

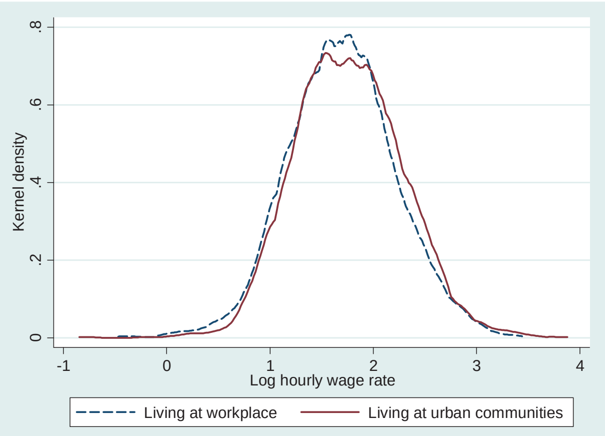


Figure 5.1: Kernel density estimates of log wage distributions, residence subsamples

In addition, as shown in Table 5.1 migrants not living at workplaces are on average more educated, older, and more experienced than their counterparts living at workplaces. A set of t-tests on the equality of means in the core human capital variables between the two subsamples further shows that their differences in education and age are significant at 1% level, though the difference is insignificant for the off-farm experience variable (p-value is 0.1552).

5.3.2.2 Migrants' distribution by sector and occupation

The RUMiC data distinguish 28 different sectors and 23 different types of occupation. Because the number of migrant workers in some of these sectors and in some occupations is very small, and because the characteristics of some sectors (occupations) are very similar, we aggregate the sectors into 14 different sectors and the occupations into 6 broad categories of occupations (see Appendix Tables 5.A1 and 5.A2 for details).

Table 5.2: Migrants' distribution by sector and occupation

	Whole sample		Female		Male	
	Obs.	%	Obs.	%	Obs.	%
Sector						
Manufacturing	1271	(19.71)	485	(19.36)	786	(19.93)
Construction	670	(10.39)	94	(3.75)	576	(14.61)
Transport and Communication	244	(3.78)	42	(1.68)	202	(5.12)
Wholesale and Retail	1697	(26.32)	786	(31.38)	911	(23.10)
Hotel and Catering Services	1174	(18.21)	562	(22.44)	612	(15.52)
Finance and Law	16	(0.25)	4	(0.16)	12	(0.30)
Real Estate	199	(3.09)	45	(1.8)	154	(3.91)
Leasing and Business Services	50	(0.78)	31	(1.24)	19	(0.48)
Scientific Research, Technical Service	195	(3.02)	36	(1.44)	159	(4.03)
Public Facilities Management	25	(0.39)	7	(0.28)	18	(0.46)
Services: Social & Household	629	(9.75)	266	(10.62)	363	(9.21)
Education, Health and Social welfare	173	(2.68)	102	(4.07)	71	(1.80)
Entertainment	80	(1.24)	36	(1.44)	44	(1.12)
Others	25	(0.39)	9	(0.36)	16	(0.41)
Occupation						
Managers or Professionals	149	(2.31)	65	(2.59)	84	(2.13)
Clerks	304	(4.71)	171	(6.83)	133	(3.37)
Sales	1016	(15.76)	567	(22.63)	449	(11.39)
Service provider	2142	(33.22)	836	(33.37)	1306	(33.12)
Production/Transportation Worker	1918	(29.75)	523	(20.88)	1395	(35.38)
Private business, self-employed	919	(14.25)	343	(13.69)	576	(14.61)

Source: Calculated from RUMiC2008.

Table 5.2 displays the distributions of migrants by sector and occupation. The industry in which migrants' work most frequently is the wholesale and retail sector (26%), followed by the manufacturing sector (20%) and the hotel and catering services sector (18%). Together these three industries account for almost two-thirds of migrant employment. For female migrants, the three sectors even account for 73% of employment. The proportion of migrants working in construction and in the social and household service sectors are close to 10% each. There is, not surprisingly, a major gender difference in working in the construction industry: construction accounts for 15% of male migrant employment and only 4% of female migrant employment. On the other hand, female migrant workers more often work in wholesale and retail and in hotel and catering services than male migrant workers.

As regards occupations, around 14% of migrants in the sample are self-employed or owners of small private businesses, with little difference between the two sexes (14% for females and 15% for males). Most migrant wage earners are employed as blue-collar workers (sales, service providers, or production/transportation workers). The share of while-collar workers (managers, professionals or clerks) among wage earners is only 8%; it is significantly larger for female migrant wage earners (11%) than for males (6%). Among the blue-collar workers, significantly more female migrants than male migrants work as a sales person, while male migrants work significantly more often as production/ transportation workers.

Figures 5.2A and 5.2B depict the sector and occupation distributions for the two residence type subsamples. Note that, unlike Table 5.2, self-employed migrants are excluded from both subsamples, as they do not report their living place in the survey data. Figure 5.2A shows there are sizeable differences in sector distribution between migrants living at workplaces and those living in urban communities. Migrants living at workplaces are most likely to work in the manufacturing sector (26%), hotel and catering services (22%), and construction (18%). Migrants living in urban communities are most likely to work in wholesale and retail trade (28%), following by manufacturing (18%) and hotel and catering services (15%). In terms of occupations, the main difference is that migrants living in urban communities are much more likely to work in a sales occupation (25%) than migrants living at workplaces (8%) – the latter are more likely to work as service producer or as production/transportation worker.

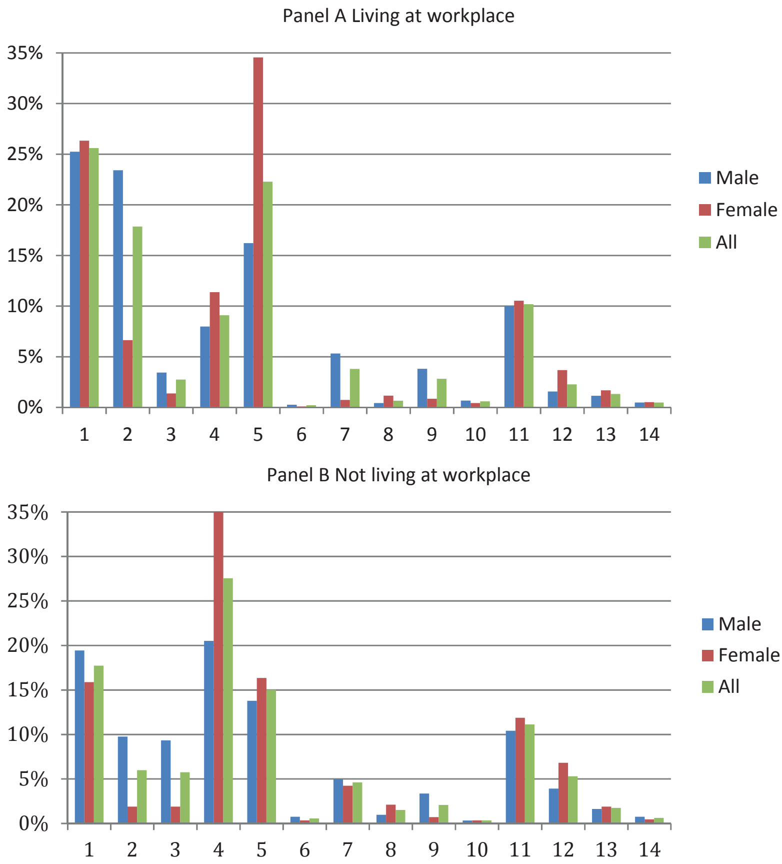


Figure 5.2A: Migrants' distribution by sector

Notes: 1 Manufacturing; 2 Construction; 3 Transport and Communication; 4 Wholesale and Retail; 5 Hotel and Catering Services; 6 Finance and Law; 7 Real Estate; 8 Leasing and Business Services; 9 Scientific Research, Technical Service; 10 Public Facilities Management; 11 Services: Social & Household; 12 Education, Health and Social welfare, 13 Entertainment;14 Others

Figure 5.2B: Migrants' distribution by occupation

Notes: 1 Managers or Professionals; 2 Clerks; 3 Sales; 4 Service provider; 5 Production/Transportation Worker

5.3.2.3 Wage rates by sector and occupation

Table 5.3 reports hourly wages by sector and occupation and their gender differences. Across all sectors, hourly wages are highest in the leasing and business services sector and in the finance and law sector. Manufacturing and construction also receive relatively higher wages than the other sectors. The wage rate in hotel and catering services, which employs 22% of female migrants, is the lowest across all sectors except for the "others" sector..

Female hourly wages are lower than male hourly wages in each sector. For the three major sectors of migrant employment (manufacturing, wholesale and retail, and hotel and catering services) the gender gap in wage rates is statistically significant. The gender differences are also economically large, more than 1 yuan per hour. The scientific research and technical service sector, employing about 3% of migrants, also exhibits an economically large and statistically significant gender wage differential.

Table 5.3: Hourly wage rates by sector and occupation

	Whole sample						Female	Male	Gender difference	
	Mean	SD	CV	p10	p50	p90	Mean	Mean	Mean	t -statistic
Sector										
Manufacturing	7.44	4.10	0.55	3.57	6.70	12.5	6.81	7.83	-1.02***	(-4.32)
Construction	7.46	5.12	0.69	3.49	6.25	12.5	6.78	7.57	-0.78	(-1.38)
Transport and Communication	7.22	4.23	0.59	3.21	6.25	12.5	6.18	7.44	-1.26	(-1.76)
Wholesale and Retail	6.71	6.16	0.92	2.5	5.21	11.9	6.00	7.32	-1.33***	(-4.44)
Hotel and Catering Services	5.96	5.01	0.84	2.68	4.76	10.42	5.28	6.59	-1.31***	(-4.51)
Finance and Law	8.68	7.92	0.91	2.98	6.25	13.02	4.14	10.20	-6.05	(-1.36)
Real Estate	6.93	3.98	0.57	2.92	5.56	13.02	5.99	7.21	-1.22	(-1.81)
Leasing and Business Services	8.73	4.26	0.49	3.99	7.81	13.13	7.86	10.14	-2.28	(-1.88)
Scientific Research, Technical Service	6.77	4.69	0.69	2.66	5.36	12.5	4.41	7.31	-2.90***	(-3.44)
Public Facilities Management	6.22	2.81	0.45	2.46	6.25	10.00	5.70	6.42	-0.71	(-0.56)
Services: Social & Household	6.52	4.7	0.72	2.68	5.36	10.42	6.52	6.52	-0.01	(-0.02)
Education, Health and Social welfare	6.12	4.33	0.71	2.5	4.76	11.90	5.64	6.81	-1.18	(-1.77)
Entertainment	6.54	3.91	0.6	2.92	5.21	11.16	6.26	6.78	-0.52	(-0.58)
Others	5.09	2.28	0.45	2.08	5.00	8.48	3.90	5.76	-1.86*	(-2.09)
Occupation										
Managers or Professionals	9.87	6.26	0.63	4.02	8.04	19.64	8.66	10.82	-2.16*	(-2.11)
Clerks	7.78	4.33	0.56	3.47	6.72	13.39	6.90	8.92	-2.02***	(-4.13)
Sales	6.11	4.91	0.8	2.68	5.21	10.42	5.64	6.71	-1.07***	(-3.48)
Service provider	5.74	3.55	0.62	2.68	4.92	9.38	5.16	6.11	-0.95***	(-6.10)
Production/Transportation Worker	7.31	3.79	0.52	3.57	6.51	12.5	6.63	7.57	-0.94***	(-4.84)
Private business owner, self-employed	8.14	8.73	1.07	2.38	5.95	14.88	7.17	8.72	-1.55**	(-2.60)

Source: Calculated from RUMiC2008. p10, p50 and p90 represent the 10, 50 and 90 ths percentiles. *, * *, *** denote significance levels of 10%, 5%, 1%, respectively.

The evidence on wage rates by occupation suggests that the best paying occupation is that of manager or professional. Clerks earn less but they earn more than blue-collar workers. Among the blue-collar workers, production/transportation workers earn most. Gender differences in all occupations are statistically significant;

they tend to be larger in the white-collar occupations. Finally, self-employed migrants (or small business owners) earn more than migrants in the other occupations except managers or professionals. Females earn on average 1.55 yuan per hour (significant at 5%) less than males among self-employed migrants.

5.3.2.4 Gender and sector/occupation wage rates

Figure 5.3 plots the correlation between the average wage rate of each industry and the share of migrant workers in a sector that is female. The size of each circle indicates the number of female migrants employed in that industry. As shown in the figure, the hourly wage rate tends to decrease as the share of females in a sector increases. For instance, the hotel and catering services sector, i.e., the second largest female employment sector, has the second lowest average hourly wage rate. The evidence suggests that the sorting of female migrants into lower-paying sectors could be an important factor in explaining migrant gender wage gaps.

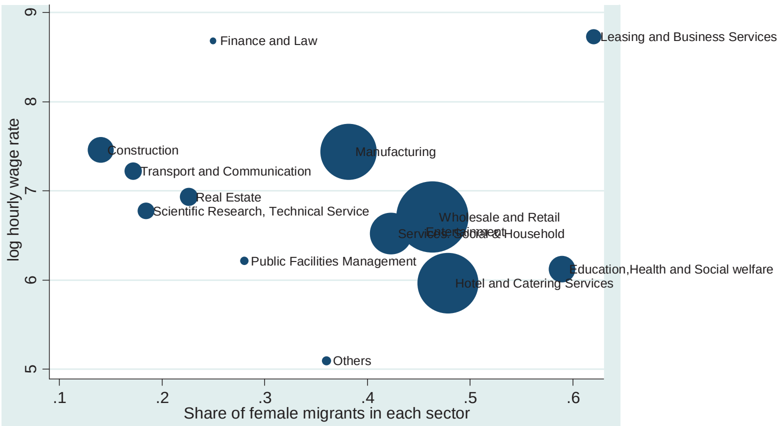


Figure 5.3: Hourly wage and female employment share by industry

Notes: Hourly wage rates on the vertical axes are calculated as the simple mean of individual observations on wages earned by migrant workers in each industry. The size of the circle represents the number of females (panel A) or males (panel B) working in the corresponding industries.
Source: Calculated from RUMiC2008.

We conduct a similar analysis for the average wage rate per type of occupation in Figure 5.4. Here the relationship is less straightforward: the share of female migrants in a specific occupation shows no clear association with the average wage rate earned in that occupation. There is thus no evidence of sorting of females into lower- or better-paying occupations. It may further be noted that the relatively well-paying occupations, i.e. managers or professionals and self-employment, employ only small shares of the total migrant population. Large numbers of migrants, both male and female, cluster in the lower-paying occupations like service jobs (see also Table 5.2).

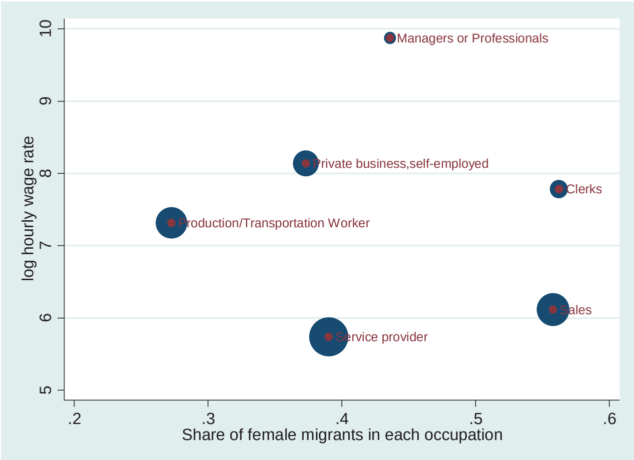


Figure 5.4: Hourly wage and female employment share by occupation

Notes: The vertical axis, hourly wage rates are calculated as the simply average of individual observations by occupations, including both male and female migrants. Data: RUMiC2008. The size of the circle represents the number of female (panel A) or male (panel B) observations in corresponding occupations.
Source: Calculated from RUMiC2008.

5.4 Empirical methodology

In this section we discuss two methods for (gender) wage gap decomposition analysis that have been used in the literature. One is the newly developed Gelbach decomposition based on a pooled wage regression strategy. The other is the well-known Blinder-Oaxaca (B-O) decomposition method. Below we provide the details about how we apply the two methods.

5.4.1 Pooled wage regressions and the Gelbach decomposition

One intuitive way to measure the unexplained gender wage gap is simply estimating a Mincer-type wage model for the pooled sample of men and women with gender dummy and other productive characteristics as explanatory variables[45]. Specifically,

$$\ln(w_i) \quad = \delta + \alpha \cdot F_i + \beta \cdot X_i + \varepsilon_i \qquad , \tag{1}$$

where the subscript i denotes individuals, ln(w) is the natural logarithm of the observed hourly wages,[46] δ is a constant, and ε is an error term, capturing all the unobserved errors. F is a dummy variable equal to 1 if the individual is a female migrant, and α is its coefficient; X is a vector of productive characteristics that determine wages, and β is the vector of related coefficients.

We consider two sets of X following Altonji and Blank (1999). One includes only education and experience measures, in line with the classic Mincer wage equation.[47] In particular, we include years of schooling, age and the quadratic term of age,[48] off-farm working experience (number of years since the first time the migrant found a job in urban area), and city dummies.[49] In the other set we further add dummies for the worker's industry and occupation. It is, however, debatable whether industry and occupation categories should be included (see Albrecht et al. 2003; Magnani and Zhu 2010; Nordman et al. 2011). Industry and occupation choices can be viewed as either the outcomes of productive characteristics or the outcomes of employer practices. Although these arguably endogenous variables may be jointly determined with wages, Albrecht et al. (2003) point out that they may reflect unmeasured human capital and may help explain wage differentials as an accounting exercise.

The parameter α—the coefficient of the female dummy—is of our primary interest.

---

[45] See Elder et al. (2010) and Gelbach (2016) for detailed discussions on the method.

[46] The use of hourly wages allows us to compare our estimates with the results from studies closest to ours, i.e., Magnani and. Zhu (2010) among others. Using weekly, monthly or annual earnings is less common in prior studies, and could embed labor supply responses (labor-leisure tradeoff) to changes in hourly wage rate (see Lemieux 2006a).

[47] Lemieux (2006b) provides detailed discussions on the specification issues of the Mincer equation.

[48] Experience is also often measured as age - schooling years - 6 (6 = compulsory school entry age). For the sake of comparability of our estimates with prior studies on China's migrants' labor market, we use age directly as a proxy for experience. This choice has a negligible impact on the empirical results.

[49] Our data set also allows us to further control for marital status and number of children, but their inclusion has negligible effects on the empirical results. It should be noted that potential omitted confounding factors such as unobservable ability or gender heterogeneity of job preferences, may still be present.

It represents how much the wage rates of females differ from observably identical males. In other words, α is a measure of the wage difference between female and male migrants that share the same characteristics, but only differ in gender. Therefore, a significant negative estimate for α would suggest that females are being discriminated.

An unbiased estimate of α requires the following condition to hold:

$$\mathrm{Cov}\ (\mathrm{F_i}, \varepsilon_\mathrm{i}|\mathrm{X_i}) = 0 \qquad (2)$$

This orthogonality assumption is hard to test, and it is easy to think of endogeneity issues which will violate the assumption. One commonly acknowledged problem, that applies equally to the estimate of β, is sample selection bias: women and men who choose to work may be different from the unemployed or inactive. In principle the Heckman two-step procedure can resolve this issue, but application of the Heckman selection model faces two difficulties in practice. Firstly, a lack of observations of unemployed and inactive individuals[50] reduces the feasibility and efficiency of first-stage regressions. Secondly, identification of the Heckman selection model requires at least one variable that affects the probability of being employed but is not directly related to an individual's wage rate. It is hard, if not impossible, to find such variables[51]. As a result we keep the sample selection issues aside as most studies in this field do (e.g., Nordman, et al. 2011; Magnani and Zhu, 2010).

Another problem, particularly relevant for the estimate of α, is the possibility of unobserved gender differences in preferences. For example, it has been reported that women are less competitive and therefore less likely to negotiate promotion into better-paying occupations or to bargain higher wages within the same occupation (see Bertrand 2011 for a review of the literature on gender differences in psychological attributes). Though there is little consensus on the existence and importance of such

[50] It is particular the case when using migrant data in China: migrants are highly mobile, e.g., searching/finding jobs elsewhere or returning to home villages once unemployed. Consequently, migrant surveys conducted in cities have few unemployed respondents. In the RUMiC 2008 sample, for instance, the number of unemployed adults is only 76 (among them, 54 are female).

[51] Lee (2012) uses age squared exclusively in the first stage regression, yielding an insignificant inverse Mills ratio.

gender differences, one should keep in mind the potential omitted variable bias when interpreting α.

Usually, a series of specifications of equation (1) with sequentially extended covariates X is estimated. The strategy, straightforward and easy-to-implement, aims to assess the contribution of additional covariates to the change of α as compared to a base specification.[52] However, Gelbach (2016) points out that a direct comparison of estimates for α from two specifications (partial vs. full) may yield inappropriate accounting for the contribution of each set of variables. He proposes the following way to disentangle the contribution of each (excluded) variable to the change in the estimate of α. Specifically, consider a full specification:

$$y \quad = X_1\beta_1 + X_2\beta_2 + \epsilon \qquad , \qquad (3)$$

with a base specification which omits the set of repressor X_2:

$$y \quad = X_1\beta_1 + \epsilon \qquad , \qquad (4)$$

where X_1 is the variable under scrutiny (female dummy in our case), X_2 contains the excluded variables (years of schooling, age, age^2, off-farm experience, industry-, occupation-, and city-fixed effects in our case). Gelbach (2016) shows that

$$\widehat{\beta_1^{base}} = \widehat{\beta_1^{full}} + (X_1{}'X_1)^{-1}X_1{}'X_2\widehat{\beta_2} \qquad (5)$$

Using equation (5), it is possible to decompose the contribution of covariate k in X_2 to $\widehat{\beta_1^{base}}$ as $(X_1{}'X_1)^{-1}X_1{}'X_{2k}\widehat{\beta_{2k}}$, where X_{2k} is column k in X_2 and is the associated coefficient for X_{2k} in the regression on y. This decomposition is thus conditioned on all other covariates and is invariant to the order in which covariates are considered (see the example in Gelbach, 2016). We use the Gelbach approach to attribute the changes in the estimated coefficient for the female dummy to different covariates.[53]

[52] This branch of literature also uses the changes in the R^2 to derive the explanatory power of included covariates (see Dickens and Katz, 1987 and Meng, 2012 for details).

[53] For Stata users, this is carried out by the user written command—b1x2 (see Gelbach, 2014). Grove, Hussey, and Jetter (2011), Buckles and Hungerman (2013), and Scholz and Sicinski (2015) are examples of recent applications of the approach.

5.4.2 Gender-specific wage regressions and the Blinder-Oaxaca decomposition

Most existing studies of factors driving gender differences in wages use the B-O decomposition approach. Appendix C explains the detailed procedures to apply the method, using three popular choices of reference wage structures. Clearly, the three approaches may yield different results and thus lead to different conclusions.[54]

Oaxaca and Ransom (1994) argue that using the male wage structure (equation (7a) in Appendix C) and the female wage structure (equation (7b)) provides the upper and lower bounds of sources of the gender wage gap. Decomposition results based on the male wage structure may be preferred for examining sources of gender wage gaps, as they show what female wage rates would be if women were paid according to the wage structure of males.

Oaxaca and Ransom (1994) further note that the wage structure under non-discrimination environments should be derived from pooled regression coefficients (i.e. the Neumark method specified in equation (8)). However, Jann (2008) and Elder et al. (2010) point out that the Neumark approach tends to overstate the contribution of productive characteristics to observed wage gaps. This is mainly due to the absence of a group dummy variable in pooled regressions. Jann (2008) thus suggests including the group dummy variable in the pooled wage regression model to derive the reference wage structure estimates in equation (8).

Taking the above literature into account, we report results for the B-O decomposition analysis based on both the male wage structure and the pooled wage structure estimated with the gender group dummy variable.

5.5 Results and discussions

5.5.1 Pooled regression analyses

5.5.1.1 Baseline results for gender wage gaps

The first question of our interest is to what extent raw gender wage differentials

---

[54] A few other choices of reference wage structures are discussed in Oaxaca and Ransom (1994) and Jann (2008). For example, Reimers (1983) suggests using the simple average coefficients from both groups' estimates, whereas Cotton (1988) proposes using group size weighted coefficients.

documented in the descriptive analysis (Table 5.1) are affected by differences in worker characteristics. Table 5.4 provides the pooled (i.e., both male and female migrant workers) regression results when different observed covariates are included (equation (1)).

Table 5.4: Pooled OLS estimates of log hourly wage rate

	(1)	(2)	(3)	(4)	(5)	(6)
	Whole sample			Wage earner		
Female=1	-0.162***	-0.153***	-0.135***	-0.168***	-0.144***	-0.118***
	(0.027)	(0.020)	(0.020)	(0.025)	(0.020)	(0.019)
Years of Schooling		0.047***	0.043***		0.048***	0.042***
		(0.004)	(0.004)		(0.004)	(0.004)
Age		0.056***	0.050***		0.058***	0.053***
		(0.007)	(0.007)		(0.004)	(0.005)
$Age^2/100$		-0.084***	-0.076***		-0.087***	-0.081***
		(0.010)	(0.010)		(0.006)	(0.007)
Off-farm experience		0.017***	0.015***		0.023***	0.021***
		(0.003)	(0.003)		(0.003)	(0.003)
Constant	1.804***	0.649***	0.923***	1.786***	0.566***	0.824***
	(0.052)	(0.128)	(0.137)	(0.058)	(0.089)	(0.102)
City FE	No	Yes	Yes	No	Yes	Yes
Industry FE	No	No	Yes	No	No	Yes
Occupation FE	No	No	Yes	No	No	Yes
Observations	6,448	6,448	6,448	4,637	4,637	4,637
R-squared	0.019	0.223	0.262	0.025	0.313	0.363

Note: Standard errors are in parentheses and are clustered at city levels. *, * *, *** denote significance levels of 10%, 5%, 1%, respectively.

Columns (1)-(3) present estimation results for the whole sample. In the first regression, as shown in column (1), we include only the female dummy. The coefficient estimate of the female dummy variable is -0.162, indicating that the raw gender wage gap equals about 16.2 percent.

In column (2) we add human capital characteristics, i.e., education, age and off-farm experience variables, together with city fixed effects. The estimate value for the wage gap changes to -0.153. Thus, the gender wage gap becomes slightly smaller when standard human capital traits and inter-city wage differentials are taken into account. Including only city fixed effects (not shown) results in a gender wage gap of -0.18, so it is the human capital variables that account for the declining gap in column (2). The estimated coefficients for human capital attributes are consistent with the existing literature. In particular, the return to schooling is highly significant and its value is comparable to the 4.1% - 4.2% return for rural-urban migrants estimated by Magnani and Zhu (2012). The coefficients for age, age squared and off-farm experience are highly significant and all have the expected signs too. These results show that a small part of the wage gap can be explained by the fact that female migrants are younger and have less schooling and off-farm experience than their male counterparts (see Table 5.1).

In column (3) we further add the industry and occupation dummies. The coefficient estimate for the female dummy becomes -0.135. Therefore, the sorting of female migrant workers into lower-paying sectors documented in Section 5.3.2.4 explains part of the wage gap, but a large unexplained gender wage gap still exists. However, as explained in the previous section, gender differences in industry and occupation may only partially be attributed to discrimination by employers. The estimates presented in columns (2) and (3) can be viewed as the upper and lower bounds of residual gender wage gaps among migrant workers respectively.

Following the previous literature, we replicate the analyses with wage earners only in columns (4)-(6). The motivation is that theoretically self-employed migrants are not directly subject to employer discrimination. The raw gender wage gap is very similar compared to the full sample, while human capital characteristics and industry and occupation sorting account for a slightly larger part of the raw gap in the sub-sample of wage earners. Including all control variables, the estimated female coefficient is still -0.118.

5.5.1.2 Differences in gender wage gaps by the type of residence

In Table 5.5, we examine the gender wage gaps for subsamples of migrant workers living at their workplace and those living in urban communities[55]. The raw gender gap is about 18 log points for both groups (see columns (1) and (4)). As additional control variables are included, the wage gap decreases more in the sample of migrants living at their workplace; according to the full model specifications (columns (3) and (6)), the

[55] Self-employed migrants are excluded for reasons discussed in Section 5.3.1.

residual wage gap for the group living at their workplace is 0.11, versus 0.16 for the group living in urban communities. Thus, the control variables account for 41% of the raw gender wage gap in the subsample living at workplaces, compared to 14% for the subsample living in urban communities.

Table 5.5: Pooled OLS estimates log hourly wage rate by the type of residence

	(1)	(2)	(3)	(4)	(5)	(6)
	Living at workplace			Not living at workplace		
Female=1	-0.183***	-0.144***	-0.108***	-0.180***	-0.161***	-0.155***
	(0.037)	(0.026)	(0.019)	(0.033)	(0.028)	(0.030)
Years of Schooling		0.042***	0.039***		0.054***	0.043***
		(0.004)	(0.004)		(0.005)	(0.006)
Age		0.059***	0.052***		0.055***	0.050***
		(0.005)	(0.006)		(0.009)	(0.009)
$Age^2/100$		-0.088***	-0.080***		-0.082***	-0.075***
		(0.007)	(0.007)		(0.013)	(0.012)
Off-farm experience		0.024***	0.020***		0.024***	0.024***
		(0.003)	(0.003)		(0.005)	(0.005)
Constant	1.756***	0.642***	0.939***	1.849***	0.557***	0.791***
	(0.058)	(0.101)	(0.119)	(0.062)	(0.154)	(0.152)
City FE	No	Yes	Yes	No	Yes	Yes
Industry FE	No	No	Yes	No	No	Yes
Occupation FE	No	No	Yes	No	No	Yes
Observations	2,866	2,866	2,866	1,771	1,771	1,771
R-squared	0.028	0.321	0.391	0.030	0.314	0.351

Note: Standard errors are in parentheses and are clustered at city levels. *, * *, *** denote significance levels of 10%, 5%, 1%, respectively.

The inclusion of industry and occupation dummies accounts for a larger reduction in the gender wage gap among migrants living at their workplace compared to migrants living in urban communities. As we can note from Figure 5.2A, industrial sorting differs considerably between the two subsamples. A more detailed analysis of the contribution of industry and occupation sorting is presented in Section 5.5.3.

5.5.2 Gender-specific regression analyses

In Table 5.6, we present the results of separate regressions for female and male migrants for different (sub-)samples. This allows us to examine to what extent the effect of productive characteristics on migrant wages differs between men and women and between their places of residence. For simplicity, we only consider the full model specification, i.e., both human capital attributes and industry and occupation dummies are controlled for.

Table 5.6: OLS estimates of log wage rate: gender specific estimates

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Whole sample		Wage earner		Living at workplace		Not Living at workplace	
	Female	Male	Female	Male	Female	Male	Female	Male
Years of Schooling	0.046***	0.039***	0.047***	0.038***	0.043***	0.036***	0.049***	0.039***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	-0.006	-0.009
Age	0.031***	0.061***	0.026***	0.068***	0.023**	0.063***	0.019	0.078***
	(0.006)	(0.009)	(0.006)	(0.007)	(0.009)	(0.007)	-0.011	-0.013
$Age^2/100$	-0.052***	-0.091***	-0.040***	-0.101***	-0.036**	-0.095***	-0.032*	-0.111***
	(0.009)	(0.013)	(0.009)	(0.010)	(0.012)	(0.010)	-0.017	-0.018
Off-farm experience	0.015***	0.015***	0.026***	0.019***	0.023***	0.019***	0.032***	0.020***
	(0.005)	(0.002)	(0.004)	(0.003)	(0.007)	(0.003)	-0.005	-0.006
Constant	1.006***	0.807***	0.992***	0.693***	1.132***	0.837***	1.048***	0.415
	(0.115)	(0.173)	(0.098)	(0.128)	(0.153)	(0.140)	-0.188	-0.244
City FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Occupation FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,505	3,943	1,799	2,838	949	1917	850	921
R-squared	0.291	0.238	0.361	0.358	0.399	0.381	0.347	0.367

Note: Standard errors are in parentheses and are clustered at city levels. *, * *, *** denote significance levels of 10%, 5%, 1%, respectively.

The first noteworthy finding is that female migrants enjoy higher rates of return to additional schooling years than males. Such gender differences holds for all samples. This finding differs from that documented in Magnani and Zhu (2012), where education return rates are 4% for both female and male migrants. It is also worth noting that the return rates to education for migrants living at workplaces are among the smallest. Secondly, the returns to age are higher for men and the age profile is a bit more pronounced for males than for females, with wages increasing more steeply until a peak level at around age 30-35 for males (slightly earlier for females) and declining more at older ages. This finding applies to all of the samples. Thirdly, the returns to additional years of off-farm experience are the same for men and women in the full sample, but are larger for females than for males for the wage earners sample — especially for the sub-sample not living at workplaces (0.32 for females vs. 0.20 for males).

5.5.3 Decomposition results

In this section we examine the contribution of different (categories of) variables to the raw gender wage gap. As described in Section 5.4, we perform both the Gelbach decomposition and the productive B-O decompositions.

5.5.3.1 Gelbach decomposition

Table 5.7 shows the results of using the Gelbach decomposition. The Gelbach approach focuses on the contributions of endowments and only allows for a detailed decomposition into the different observed characteristics. A positive sign in the column "contribution" indicates that the relevant variable contributes to a reduction in the gender wage gap.

The results reveal several important findings. First of all, all covariates combined account for only 16.6% of the raw gender wage gap in the full sample, and almost 30% in the sample of wage earners. For the sample living at workplaces the observed characteristics even account for more than 40%. By contrast, they account for only 13% of the raw gender wage gap among migrants living in urban communities.

Table 5.7: Gelbach decomposition of gender wage gaps

	Whole sample		Wage earner		Living at workplace		Living in urban community	
	contribution	% of gap	contribution	% of gap	contribution	% of gap	contribution	% of gap
Schooling Years	-0.011***	6.9%	-0.007**	4.2%	-0.007*	3.8%	-0.012**	6.7%
Age & Age2	0.005**	-3.3%	0.004	-2.4%	0.002	-1.1%	0.001	-0.6%
Off-farm experience	-0.012***	7.3%	-0.021***	12.5%	-0.019***	10.4%	-0.030***	16.7%
Industry	-0.021***	13.0%	-0.027***	16.1%	-0.040***	21.9%	-0.003	1.7%
Occupation	-0.003	2.0%	-0.006	3.6%	-0.008	4.4%	-0.006	3.3%
City	0.015***	-9.3%	0.009	-5.4%	-0.003	1.6%	0.026**	-14.4%
Total	-0.027***	16.6%	-0.050***	29.8%	-0.075***	41.0%	-0.024	13.3%
Raw Gap	-0.162***	100%	-0.168***	100%	-0.183***	100%	-0.180***	100%

Note: *, * *, *** denote significance levels of 10%, 5%, 1%, respectively.

Clearly, such differences in total contributions are due to the differences in the specific contributing factors among different samples. Most importantly, for the subsample living at workplaces, the major contributing factor is the industry fixed effects, whereas the contributions of city fixed effects and age are negligible. It thus suggests that the inter-industry wage differentials and gender differences in industry sorting are the major contributor to the gender wage gap in this sample. From Figure 5.2A we see that a third of the female migrants living at their workplace are employed in the hotel and catering services sector, which is the lowest paying sector (Table 5.3). Male migrants living at their workplace, on the other hand, are much more likely to work in construction.

In the sample of migrants living in urban communities, the city fixed effects show a large contribution towards reducing the gender wage gap, while industry fixed effects contribute very little. This finding suggests that females that do not live at their workplaces tend to work in better-paying cities than men while gender differences in industry sorting are much less relevant for this sub-sample.

Combining the two samples (the columns for wage earners in Table 5.8), we find that the industry fixed effects dominate the city fixed effects. It implies that the effect of sorting of females living at their workplaces into lower-paying sectors is more pronounced than the positive effect of working in better-paying cities by females living in urban communities.

Apart from the above results, the standard human capital characteristics play similar roles in explaining the gender wage gap compared to findings in previous studies. The contribution of the difference in schooling between male and female migrants is small: 6.5% for migrants living in urban communities and 3.8% for those living at their workplaces. The difference in off-farm experience between the two sexes contributes 16.7% to the wage gap of migrants living in urban communities and 10.4% to that of migrants living at their workplaces. The variable age enlarges gender wage gaps in both groups, but only has a small contribution.

Altogether, we find that there are considerable differences in contributing factors stemming from the sample being considered. Most notably, human capital variables play a relatively limited role in explaining the gender wage gap among migrants living at workplaces, while industry sorting is the main contributor for this group. Among migrants living in urban communities, however, industry sorting has hardly any impact while schooling and off-farm experience contribute more to the gender wage gap. These findings imply that the role of gender differences in characteristics (mainly driven by differences in industry sorting) is likely to be underestimated in studies using survey data of Chinese migrants that do not cover the migrants living at workplaces. Conversely, the contribution of the coefficient effect - the "unexplained" part in the Gelbach decomposition - is relatively small among migrants living at workplaces and is likely to be overestimated in studies using residence-based survey data.

5.5.3.2 Blinder-Oaxaca decomposition

To investigate the robustness of the decomposition results reported so far, we employ B-O decompositions with pooled and male wage structure as the reference wage structures. Table 5.8A shows the detailed decomposition results for the endowment effects, and Table 5.8B for the coefficient effects. [56]

As expected, the B-O decomposition results for the endowment effects are consistent with the results using the Gelbach method. Turning to the coefficient effects, which are not covered by the Gelbach method, we find that gender differences in returns to age account for more than 300% of the raw gender wage gap; gender differences in coefficients on all other covariates (including schooling and off-farm experience) contribute negatively – reducing the gender wage gap. Results presented in Table 5.6 showed that the estimated returns to schooling and off-farm experience are higher for female migrants than for male migrants. The decomposition results confirm this, but also show that the contributions of these “female advantages” are statistically insignificant. In fact, the main negative contribution comes from the constant term, indicating a higher conditional mean wage for women after accounting for the observed characteristics. In other words, gender differences in some unobserved characteristics - or differences in the returns to these characteristics - contribute to a narrowing of the gender wag gap in all subsamples.

[56] To deal with the so-called omitted group problem, the contributions of categorical variables have all been normalized using the method proposed by Gardeazabal and Ugidos (2004). In addition, Note that a positive contribution in the B-O decomposition analysis indicates that the corresponding variable contributes to a widening of the gender wage gap (as opposed to the interpretation in Gelbach decompositions).

Table 5.8A: Blinder-Oaxaca decomposition of gender wage gaps: the endowment effects

	(1)		(2)		(3)		(4)	
	Whole sample		Wage earner		Living at workplace		Living in urban community	
	contribution	% of gap	contribution	% of gap	contribution	% of gap	contribution	% of gap
Reference wage structure = using pooled coefficients								
Schooling Years	0.011***	6.8%	0.007*	4.2%	0.007	3.8%	0.012*	6.7%
Age & Age2	-0.005*	-3.1%	-0.004	-2.4%	-0.002	-1.1%	-0.001	-0.6%
Off-farm experience	0.012***	7.4%	0.021***	12.5%	0.019***	10.3%	0.030***	16.7%
Industry	0.021***	13.0%	0.026***	15.5%	0.040***	21.7%	0.002	1.1%
Occupation	0.003	1.9%	0.007	4.2%	0.008	4.3%	0.006	3.3%
City	-0.011*	-6.8%	-0.003	-1.8%	0.010	5.4%	-0.025**	-13.9%
Total	0.027**	16.7%	0.050***	29.8%	0.075***	40.8%	0.024	13.3%
Reference wage structure = using male coefficients								
Schooling Years	0.010***	6.2%	0.007*	4.2%	0.006	3.3%	0.011*	6.1%
Age & Age2	-0.003	-1.9%	-0.002	-1.2%	-0.000	0.0%	0.004	2.2%
Off-farm experience	0.011***	6.8%	0.019***	11.3%	0.018***	9.8%	0.024***	13.3%
Industry	0.024***	14.8%	0.032***	19.0%	0.043***	23.4%	0.011	6.1%
Occupation	-0.000	0.0%	-0.002	-1.2%	0.007	3.8%	-0.008	-4.4%
City	-0.008	-4.9%	-0.002	-1.2%	0.011	6.0%	-0.021*	-11.7%
Total	0.030**	18.5%	0.046***	27.4%	0.077***	41.8%	0.022	12.2%
Raw Gap	0.162***	100%	0.168***	100%	0.184***	100%	0.180***	100%

Note: *, * *, *** denote significance levels of 10%, 5%, 1%, respectively.

Table 5.8B: Blinder-Oaxaca decomposition of gender wage gaps: the coefficient effect

	(1)		(2)		(3)		(4)	
	Whole sample		Wage earner		Living at workplace		Living in urban community	
	contribution	% of gap	contribution	% of gap	contribution	% of gap	contribution	% of gap
Reference wage structure = using pooled coefficients								
Schooling Years	-0.060	-37.0%	-0.084	-50.0%	-0.059	-32.1%	-0.089	-49.4%
Age & Age2	0.504***	311%	0.640***	381%	0.578***	314.1%	0.973***	541%
Off-farm experience	-0.002	-1.2%	-0.026	-15.5%	-0.013	-7.1%	-0.048*	-26.7%
Industry	-0.017	-10.5%	-0.026	-15.5%	-0.050*	-27.2%	-0.015	-8.3%
Occupation	-0.029*	-17.9%	-0.026	-15.5%	0.013	7.1%	-0.062*	-34.4%
City	-0.003	-1.9%	0.001	0.6%	-0.002	-1.1%	-0.000	0.0%
_cons	-0.274	-169%	-0.375*	-223%	-0.355	-193%	-0.625*	-347%
Total	0.135***	83.3%	0.118***	70.2%	0.109***	59.2%	0.156***	86.7%
Reference wage structure = using male coefficients								
Schooling Years	-0.059	-36.4%	-0.083	-49.4%	-0.058	-31.5%	-0.087	-48.3%
Age & Age^2	0.502***	310%	0.639***	380%	0.576***	313%	0.968***	538%
Off-farm experience	-0.002	-1.2%	-0.024	-14.3%	-0.012	-6.5%	-0.043	-23.9%
Industry	-0.021	-13.0%	-0.032	-19.0%	-0.052	-28.3%	-0.025	-13.9%
Occupation	-0.026*	-16.0%	-0.017	-10.1%	0.013	7.1%	-0.048	-26.7%
City	-0.005	-3.1%	0.000	0.0%	-0.003	-1.6%	-0.004	-2.2%
_cons	-0.274	-169%	-0.375*	-223%	-0.355	-193%	-0.625*	-347%
Total	0.132	81.5%	0.122***	72.6%	0.106***	57.6%	0.158***	87.8%
Raw Gap	0.162***	100%	0.168***	100%	0.184***	100%	0.180***	100%

Note: *, * *, *** denote significance levels of 10%, 5%, 1%, respectively.

5.5.4 Gender and inter-industry wage premium

The above decomposition results have demonstrated the importance of industry sorting in accounting for gender wage gaps, especially among migrants living in workplaces. In this section, we discuss the industry-specific wage premiums to further analyze how sorting into different industries plays a role in wage determination. In Table 5.9 we report the industry coefficients; they are based on equation (1) with a full set of explanatory variables. They can be interpreted as industry-specific wages relative to the base category, i.e. the manufacturing industry, conditional on workers' productive characteristics, city of residence, and occupation.

Table 5.9: Inter-industry wage premium estimates: sub-samples by residence

	Pooled		Female		Male	
	Estimate	SD	Estimate	SD	Estimate	SD
Living at workplace						
Construction	0.047	0.037	0.009	0.071	0.035	0.040
Transport and Communication	0.080	0.050	-0.098	0.062	0.091	0.057
Wholesale and Retail	-0.094	0.070	-0.071	0.085	-0.120	0.072
Hotel and Catering Services	-0.160	0.047***	-0.146	0.050 **	-0.186	0.055***
Finance and Law	-0.001	0.122	-0.612	0.072 ***	0.092	0.158
Real Estate	-0.018	0.072	-0.014	0.206	-0.042	0.071
Leasing and Business Services	0.186	0.083**	0.138	0.091	0.275	0.121**
Scientific Research, Technical Service	-0.128	0.031***	-0.041	0.136	-0.156	0.034***
Public Facilities Management	0.054	0.055	0.140	0.138	0.027	0.064
Services: Social & Household	-0.114	0.065	-0.027	0.079	-0.168	0.071**
Education, Health and Social welfare	-0.157	0.093	-0.149	0.119	-0.171	0.119
Entertainment	0.061	0.084	0.099	0.093	0.025	0.096
Other	-0.178	0.049***	-0.147	0.112	-0.236	0.062***
Not living at workplace						
Construction	0.003	0.058	0.043	0.088	0.005	0.076
Transport and Communication	-0.136	0.058**	-0.047	0.056	-0.149	0.082*
Wholesale and Retail	-0.043	0.050	0.010	0.063	-0.093	0.059
Hotel and Catering Services	-0.040	0.057	-0.064	0.067	0.010	0.083
Finance and Law	0.194	0.136	0.102	0.179	0.154	0.226
Real Estate	0.051	0.060	0.038	0.069	0.070	0.094
Leasing and Business Services	0.114	0.101	0.259	0.081 ***	-0.122	0.180
Scientific Research, Technical Service	0.064	0.103	-0.353	0.170 *	0.132	0.093
Public Facilities Management	-0.094	0.119	-0.046	0.208	-0.136	0.081
Services: Social & Household	0.017	0.067	0.114	0.059 *	-0.059	0.091
Education, Health and Social welfare	-0.184	0.074**	-0.141	0.077 *	-0.193	0.090*
Entertainment	-0.051	0.183	-0.005	0.211	-0.114	0.192
Other	-0.383	0.158**	-0.352	0.214	-0.462	0.240*

Notes: Reference industry = Manufacturing. Full sets of explanatory variables, i.e., female dummy (when applicable), years of schooling, age, age2, off-farm experience, and industry-, occupation, and city- fixed effects, are included. Standard errors are clustered at city levels. *, * *, *** denote significance levels of 10%, 5%, 1%, respectively.

As can be seen from the table, the leasing and business services sector is the best-paying industry: estimates for this sector are positive and generally larger than other sectors. This sector, however, employs very few migrants; it therefore contributes little to explaining gender wage gaps among migrants. Most industries have negative coefficients, indicating that jobs in manufacturing (the reference industry) are relatively well-paid.

Estimates for the construction sector are generally positive, albeit insignificant. Recall that the construction industry is the most male-dominated industry, and employs a large share of migrants living at workplaces. The hotel and catering service sector, which is the most female-dominated sector in the subsample living at workplaces, has large negative and significant coefficients. In conclusion, the relatively high wages in construction and manufacturing and the relatively low wages in hotel and catering services explain to a great extent the large effect of industry sorting on the gender wage gap among migrants living in workplaces.

5.6 Conclusions

The 2010 China population census data report that over 220 million people have migrated from rural to urban areas (NBS, 2011). The females among these migrants may face a double disadvantage at the urban labor market, but rigorous analyses of gender pay differentials among rural-urban migrants in China are still scarce. Data sets used in the few available studies on this issue focus on long term migrants residing in urban communities, while largely excluding migrants living at their workplaces. In this study we investigate the gender wage gap and its contributing factors using more representative migrant data, namely nationally representative migrant data collected by the RUMiC project in 2008. The data set covers both migrants living at workplaces and migrants living in urban communities.

Our study finds a raw gender gap of about 16-18% in this data set, which is smaller than gender gaps of around 26-30% found in previous studies. It suggests that the gender wage gap among migrants was possibly declining in the years before the survey, as we did not find significant differences in raw gender wage gaps between the two sub-samples of migrants living at workplaces and those living in urban communities. Our decomposition analyses further show that a relatively small part of the gender wage gaps among migrants can be explained by endowments differences, and that the magnitude of the endowment effects differs substantially between the two groups of migrants. Empirical studies on gender wage gaps usually find that around one half of the gap can be explained by migrants' characteristics, but previous studies on rural-urban migrants in China (Magnani and Zhu, 2012; Qin et al. 2016) found smaller contributions of migrant endowments to the gender wage gap. In our study we find that migrant endowments account for around 40% of the gender wage gap among migrants living at working places and for less than 14% of the gender wage gap among migrants living in urban communities. The main sources of the gender wage differentials also differ substantially between the two sub-groups, with gender differences in industry sorting playing an important role for migrants living at their workplace and differences in education and experience being the main driving force of the gender wage gap of migrants living in urban communities.

We further find that wage rates of males grow faster with age than those of females, especially at the early stages of career, in both sub-groups. This finding is consistent with those of other studies of gender wage gaps and might point at the existence of a glass ceiling effect for women. Another noteworthy finding is that returns to additional years of schooling and off-farm experience are larger for females than for males, especially for migrants living at the workplaces. This finding suggests that increasing education and work experience among migrants will help narrow down the gender wage gap, particularly among those living at workplaces. Finally, it should be emphasized that the unexplained part is relatively large in all our decomposition analyses. In other words, gender differences in some unobserved characteristics or differences in the returns to these characteristics make up a relatively large part in the observed raw wage gaps. More research is needed to identify these contributing factors which seem to be beyond conventional factors identified in gender wage gap studies.

Appendix A Reclassification of Occupation categories

Table 5.A1: Reclassification of Occupation categories

Occupation	New code	RUMiC code (c104)	
Managers & Professionals	1	1	Professionals
		2	Managers
Clerks	2	3	Clerks
Sales	3	4	Retail of agricultural by-products
		5	Sales
		6	Recycling and other buyer
Service provider	4	8	Restaurant and hotel staff
		9	House-maids, household worker
		10	Hairdresser, beautician, Masseur, tourist guide
		11	Car and home appliance repair
		12	Cleaning and sanitizing
		13	Chefs and butcher
		14	Kitchen assistance
		15	Security, warehouse and property management
		17	Other service area
		25	Family business helper and others
Production/transportation worker	5	7	Delivery and transport worker
		16	Drivers and conductors
		18	Construction laborers
		19	Manufacturing
		20	Repair and manufacturing service
		21	Other factory process
Private business owner & Self-employed	6	22	Private business owner
		23	Self-employed

Note: Occupation categories are re-grouped based on one digit occupation classification provided by RUMiC surveys. The occupation—farmers—is eliminated (involving very few observations).

Appendix B Reclassification of Industry categories

Table 5.A2: Reclassification of Industry categories

Sector	New code	RUMiC Code (c106)	
Manufacturing	1	3	Manufacturing
		4	Production and Supply of Electricity, Gas and Water
Construction	2	5	Construction Enterprise
Transport, Communication	3	6	Transport, Storage and Post
		7	Information Transmission, Computer Services and Software
Wholesale and Retail	4	8	Wholesale and Retail Trade
Hotel and Catering Services	5	9	Hotel and Catering Services
Finance and Law	6	10	Bank
		11	Security Activities
		12	Insurance
		14	Law
Real Estate	7	13	Real Estate
Leasing and Business Services	8	15	Leasing and Business Services-Accountant
	8	16	Leasing and Business Services-Others
Scientific Research, Technical Service	9	17	Scientific Research, Technical Service
Public Facilities Management	10	18	Management of Water Conservancy, Environment and Public Facilities
Services: Social& Household	11	19	Services-Social Intermediary/agency
		20	Services-Tourist Guide
		21	Services-Others
		29	Services to Households
Education, Health and social welfare	12	22	Education
		23	Health
		24	Social Security and Social Welfare
Entertainment	13	26	Entertainment
Others	14	25	Journalism and Publishing Activities
		27	Public Management and Social Organization
		28	International Organization

Note: Agriculture (including Forestry, Animal husbandry, Fishery) and mining sectors, coded as 1 and 2 in raw RUMiC2008 data, are excluded.

Appendix C Blinder-Oaxaca Decomposition

In this Appendix, we provide an explanation to the different decomposition results can yield from different choices of reference wage structures. Suppose, Mincer-type wage functions, estimated separately for men and women, are of the following form

$$\ln(w_i^g) = \delta^g \cdot X_i^g + \mu_i^g \qquad , \tag{6}$$

where i denotes individuals, ln(w) is the natural logarithm of the observed hourly wage, X is the vector of productive characteristics (including a constant term), δ is the vector of coefficients, μ is an error term, and g denotes the group indicator for females (f) and males (m), respectively. The B-O approach decomposes the difference in mean wages between males and females as follows:

$$\overline{\ln(w^m)} - \overline{\ln(w^f)} = \delta^m\left(\overline{X^m} - \overline{X^f}\right) + (\delta^m - \delta^f)\overline{X^f} \tag{7a}$$

Hence, the mean log wage difference consists of two parts: the wage difference accounted for by different productive characteristics (the first term on the right-hand side) and the wage gap stemming from different gender-specific returns to these characteristics (the second term on the right-hand side). Different terminologies have been used to refer to the two terms: endowment effect vs. coefficient effect, quantity vs. price, explained vs. unexplained/discrimination effect, etc. We choose the terms endowment effect and coefficient effect throughout the rest of the paper.

Alternatively, the difference in mean wages may be decomposed as:

$$\overline{\ln(w^m)} - \overline{\ln(w^f)} = \delta^f\left(\overline{X^m} - \overline{X^f}\right) + (\delta^m - \delta^f)\overline{X^m} \tag{7b}$$

Here, the endowment effect is evaluated using coefficient estimates from the female sample (rather than the male sample, as in equation 7a). In other words, the reference wage structure in (7a) is the male wages structure, while in (7b) the reference wage structure is the female wage structure. Conversely, the coefficient effect in (7b) is evaluated using male characteristics, while it is evaluated using female characteristics in equation (7a).

Neumark (1988) proposes another widely used decomposition based on the so-called pooled coefficients:

$$\overline{\ln(w^m)} - \overline{\ln(w^f)} = \delta^*\left(\overline{X^m} - \overline{X^f}\right) + [(\delta^m - \delta^*)\overline{X^m} + (\delta^* - \delta^f)\overline{X^f}] \tag{8}$$

where δ^* is the vector of coefficient estimates obtained from a wage regression using the pooled sample of females and males (without a group dummy).

CHAPTER 6

Conclusion

This thesis focuses on several major challenges associated with the urbanization process in China, thereby focusing on two dimensions: 1) the development of the industrial economy, particularly the effects of the real estate booms on manufacturing development; 2) problems facing specific socio-economic groups, in particular land conflicts and gender wage inequality among migrants. To obtain a better understanding of the causes of these problems, I use a range of data sets at firm, industrial, regional, and individual/household levels to investigate the potential causes. In this chapter, I synthesize the findings in different chapters, link them to ongoing academic debates, and present recommendations for policies and future research.

6.1 Synthesis

6.1.1 Real estate booms and manufacturing development

The question whether real estate booms in China hindered the development of industrial sectors is a hot policy and academic issue in China. In Chapter 2 I examine the effects local housing price appreciations on manufacturing firms' investment rates. The empirical evidence suggests that rising housing prices can discourage manufacturing investment; the negative effect of housing price on firm investments is more pronounced for firms that are labor intensive, non-state owned, and have few linkages with real estate sectors. Similar disruptive effects of housing prices on the industrial sector are also documented in Chen et al., (2015) and Rong et al., (2016). By contrast, a study on the US suggests that real estate price appreciations have positive effects on firm investments as housing price appreciations can increase firms' debt capacity and thus investment — the so-called collateral channel (Chaney et al, 2012). In this thesis I provide evidence that housing price growth can push up input cost — the so-called Dutch disease channel. To reconcile the mixed evidence, it may be realized that theoretically housing price appreciations may increase debt capacities but not necessarily the realized investments — firms' investment decisions also depend on the potential investment opportunities. It is also possible that the geographical unit used for analyses may affect the empirical relationship between housing price and investment: the negative housing price effects on manufacturing investment may diminish as the distance to an urban center increases. Thus, stronger negative (or less positive) effects can be expected when a smaller geographical unit is used.

Next, I switch to industrial-level data and investigate if real estate booms reduce intra-industry resource allocation efficiency (Chapter 3). I find that industries intrinsically highly linked with real estate sectors experienced increasing dispersion in firm-level total factor productivity (TFP). These results indicate that the real estate boom in China reduced inter-firm competition pressure and contributed to a decline in resource allocation efficiency.

Taken together, the two chapters indicate that real estate booms in China had negative effects on manufacturing investment (at local level) and reduce intra-industry resource allocation efficiency.

6.1.2 Regional competition and land conflicts

Available studies of the Chinese economy argue that the so-called regional-competition-growth model is a fundamental factor underlying China's spectacular economic growth (see Xu, 2011 for a review of the literature). According to these studies, fiscal decentralization encourages local officials to produce desirable public goods and to adopt pro-growth policies, while political centralization empowers the central government to use political promotion to reward local economic performance (e.g., attracting FDI). Supportive empirical evidence of this theory has been documented in Li and Zhou, (2005), among others.

However, this growth model, as shown in Chapter 4, also contributes to land conflicts. I find that the cross-product of FDI and fiscal decentralization has a significant positive effect on land conflicts (as proxied by land areas involved in illegal land use cases), which indicates that the impact of FDI on land conflicts is largest when the degree of fiscal decentralization is high. Hence, the regional-competition-growth model turns out to be less of a success story when the social costs of land conflicts are taken into account.

6.1.3 Rural-urban migrants, gender, and wage gaps

Rural-urban migrants are an important group within China's labor force. According to the 2010 China population census, over 220 million rural migrant workers are living in urban areas (NBS, 2011). Due to the *Hukou* system, labor markets are segmented between rural migrants and urban residents (see Meng, 2012 for a survey).

An under-explored aspect of rural-urban migration is the difference in wages between male and female migrants. In Chapter 5, I tests whether the wage differences can be attributed to human capital characteristics, gender discrimination, or other causes. The results show that most of the gender wage gap cannot be attributed to gender differences in observed characteristics like education or sector of employment. Both migrants residing in urban communities and those living at their workplaces are included in the analysis; the latter group is largely excluded in previous studies. One important implication from the chapter is that female migrants in China's urban labor market therefore face a "double-disadvantage" – *hukou* and gender discriminations.

6.2 Discussion

I discuss in this section the link between the main findings listed above and the overarching objective of the thesis — the impact of processes related to urbanization in China on the development of the industrial economy and on problems facing specific socio-economic groups. Existing studies of urbanization indicate that problems and challenges related to market and/or policy failures, such as traffic congestion, environment pollution, and provision of public goods, need to be carefully managed (see section 1.1.1 in Chapter 1).

Distinctive features of urbanization in China and elsewhere are the conversion of land and the migration of labor from rural to urban areas. The challenges associated with land and labor urbanization in China can also be thought of in terms of market and policy failures. Specifically, firms may experience negative effects from the boom of real estate markets, e.g., rising production costs like wages. The spillovers from the housing market are not easily mitigated via, e.g., moving to a new location, as the sunk cost of an existing production line may be high; or hiring cheaper workers, as the inter-city movements of workers are also not free — for instance, workers often need to buy/sell houses for job changes, which will cost time and money. In the case of land conflicts, both market and policy failures are at work. Conversion of rural land to urban land can only be done by local governments; but local governments often base their decisions on other criteria (like government budgets or promotion) than economic efficiency and social welfare. Rural-urban migrants face policy failures through the *hukou* system, which increases the costs of labor movements. And, as shown in Chapter 5, gender differences in wage payments among migrants are not mainly driven by market forces, i.e., differences in worker characteristics. Hence, there seems also a market failure at work in the labor market of migrants.

6.3 Policy implications

The Economist (2014) wrote on China's urbanization: 'In building its cities, China's officials have had only one great idea in mind: growth. That has brought huge benefits and problems too'; 'China needs to change the way it builds and runs its cities'. Indeed, in this thesis I find evidence which suggests that China's urbanization has contributed to serious economic and social problems. As summarized in the previous section, these issues are largely due to different kinds of market and policy failures. Accordingly, addressing these problems implies addressing these failures. Overall, the policy implications arising from this thesis can be summarized as follows: addressing the negative effects of the rising housing prices on industrial development, improving the institutional environment of urbanization, and investing in equal access to urban labor markets.

6.3.1 Coordinating urbanization and industrialization

The first policy recommendation is to improve the coordination between urbanization and industrialization. China was viewed as lagging in urbanization given its industrialization and income levels in the 1990s, when the *hukou* registration system impeded urbanization by giving migrants no formal access to urban public goods (Henderson, 2009). Since the early 2000s, urbanization has been accelerating after the reforms in land and housing markets. Currently, urbanization in the third- forth- tier cities in China is facing increasing housing stocks (Glaeser, 2016). It is particularly important in these cities to better coordinate urbanization and industrialization.

According to the evidence shown in Chapters 2 and 3, the real estate boom has negative effects on the development of the industrial sector. It is thus advisable for governments to increase the supply of affordable/social houses and to consider the introduction of property tax to mitigate the negative effects of rising housing price. In addition, Chinese governments have intensively used real estate construction as an instrument to promote urbanization and thereby economic growth. This policy should be withdrawn in the future to better coordinate urbanization with industrialization.

6.3.2 Improving institutional environments of urbanization

The second policy recommendation is to improve the institutional environment of urbanization in China. Overall, urbanization operated under a government-dominated institutional environment: the rural-to-urban movements of land and labor in China are governed by the monopoly power of governments on land and the *hukou* restrictions to labor mobility, while the fiscal system allows governments to raise considerable revenues from land acquiring and leasing. Adjustments are needed in these institutional arrangements to reduce their negative effects. Specifically, rural land should either be privatized, or public land requisitions should be subject to transparent processes and fair negotiation. In addition, fiscal arrangements may be further improved by re-classifying local governments' revenues sources and expenditure responsibilities to limit their incentives to engage in land finance.

6.3.3 Investing in the equal access to urban labor market

The third policy recommendation is to invest in equal access to urban labor markets. The incomes earned by, especially female, migrants remain substantially below those of urban counterparts with similar characteristics. Full relaxation of *hukou* restrictions for migrants and allowing equal access to urban public goods for migrants and urban residents would stimulate more rural migrants to settle down in cities and benefit from, e.g.., the knowledge spillovers in cities.

Other policy options include stimulating on the job training for migrants in urban areas. Education levels of migrants are usually lower than the levels of their urban counterparts. Improving the rural education system is also likely to contribute to a better access of migrants to urban labor markets.

Overall, the supply of public goods in cities and in rural areas can help rural population have better access to urban labor markets and make the best use of China's urbanization benefits.

6.4 Recommendations for future research

The research presented in this thesis aims to improve the understanding of several emerging challenges associated with the rapid urbanization in China in recent years. It, however, may need a bundle of researches instead four essays to address them. In this section I discuss the possible extensions that can be made.

One main implication of my research is that urbanization needs to deal with the challenges of rising housing prices, or rising geographical rents in general. I have shown evidence that the boom in the real estate sector had negative effects on manufacturing investment and intra-industry resource allocation efficiency in China — from an efficiency point of view. I do not examine the possible income distribution effects of housing prices. Future research, therefore, can compensate by incorporating the possible income/wealth distribution effects of rising housing prices. In his book "Capital in the Twenty-First Century", the French economist Piketty (2014) has shown evidence that housing wealth inequality is an important source of income inequality in capitalist economies such as Britain, France, the US, and Germany since the mid-20th-century. There is, by contrast, lack of systematic evidence for developing countries like China.

Future research may further investigate the factors underlying the positive relationship between housing prices and wealth inequality, if there is one. Some studies argue that housing wealth growth coexists with rising mortgage debts, which will enter the pockets of the creditors (i.e., the rich part of the population) — and thereby contribute to inequality (see, the Economist, 2015). This answer is unsatisfying if we further ask why (rational) people would like to carry housing debts; there must be some other factors are at work. One possible answer may lie in imperfections on financial markets: creditors charge higher interest rates than can be justified by housing values gains. Of course, there may be other factors that play a role. In particular, urban economics studies have pointed out that rising geographical rents are reflections of the increasing power of locations in the information era: the development of E-commerce, for instance, enables firms to manage a global-wide market from a specific location; also, information technology spurs the demand of ideas — which is produced by working ('living') together with smart people — obviously in crowded locations (Glaeser, 2012). In this sense, 'productive space' can attain higher profit margins, i.e., geographical rents (These rents will partially become capital rents via, e.g., housing debts, and contribute to income inequality). Further research can thereby provide new explanations/evidence on possible income distribution effects of rising housing prices.

There are also a few immediate extensions that can be made to the studies presented in this thesis. Chapters 2 and 3 together investigate firms' responses to a housing boom. A natural extension of this research is to investigate how people move across locations in responses to changing local housing price levels. Chapter 4 explores causes of land conflicts in China. The exposure to land conflicts (or land acquisition in general) may have a range of impacts that deserve further investigation, e.g., trust into the legal system, attitudes towards government, etc. Chapter 5 is a preliminary examination of gender wage gaps among migrants in China. Potential extensions include: to what extent the urbanization process in China has changed the gender roles within migrant households; from a policy perspective, what are the policy options to reduce gender wage gaps in the migrant labor market?

Lastly, Chapter 3 reminds the importance of heterogeneity in practice (and in theory). Heterogeneity refers to different characteristics (productivity, preference, etc.) of micro-economic units, i.e., individuals, firms, etc. Banerjee et al. (2005), among others, emphasize that it is important to incorporate heterogeneity in growth models; urban economics and trade theories have seen important progress after accounting for worker and firm heterogeneity (Syverson 2011; Bernard et al., 2011). Research along this line is still in its infancy; this will be another fruitful area of future research.

References

Acemoglu, D., Yared, P., 2010. Political limits to globalization. American Economic Review 100 (2): 83–88.

Adelino, M., Schoar, A. & Severino, F. (2015). House prices, collateral, and self-employment. Journal of Financial Economics.

Akai, N., Sakata, M., 2002. Fiscal decentralization contributes to economic growth: evidence from state-level cross-section data for the United States. Journal of Urban Economics 52 (1): 93 – 108.

Albrecht, J., Björklund A. & Vroman S., 2003. Is there a glass ceiling in Sweden? Journal of Labor Economics 21 (1): 145–77.

Alfaro, L., Chanda, A., Kalemli-Ozcan, S. & Sayek, S., 2004. FDI and economic growth: the role of local financial markets. Journal of International Economics 64 (1): 89–112.

Alfaro, L., Chari, A., 2014. Deregulation, misallocation, and Size: evidence from India. Journal of Law and Economics 57, 897-936.

Alfaro, L., Charlton, A. & Kanczuk, F., 2008. Plant-size distribution and cross-country income differences. NBER Working Paper No. 14060. National Bureau of Economic Research, Cambridge, MA.

Altonji, J. G., Blank, R. M., 1999. Chapter 48 Race and gender in the labor market. in Ashenfelter, Orley C., Card, David (Eds.) Handbook of Labor Economics, 3, Part C:3143–3259. Elsevier.

Arellano, M., Bond, S., 1991. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. Review of Economic Studies 58 (2): 277–297.

Arellano, M., Bover, O., 1995. Another look at the instrumental variable estimation of error-components models. Journal of Econometrics 68 (1): 29 – 51.

Asiedu, E., Lien, D., 2011. Democracy, foreign direct investment and natural resources. Journal of International Economics 84 (1): 99–111.

Autor, D. H., Katz, L. F. & Kearney, M. S., 2008. Trends in US wage inequality: revising the revisionists. The Review of Economics and Statistics 90 (2): 300–323.

Banerjee, A., Duflo, E., 2005. Growth theory through the lens of development economics. In: Aghion, Philippe, Durlauf, Steven N. (Eds.), Handbook of Economic Growth, Vol. 1, Part A. Elsevier, Amsterdam, pp. 473-552.

Barro, R. J. (1976). The loan market, collateral, and rates of interest. Journal of money, Credit and Banking, 439-456.

Bartelsman, E., Haltiwanger, J. & Scarpetta, S., 2013. Cross-country differences in productivity: the role of allocation and selection. American Economic Review 103, 305-334.

Beaudry, P. & Ethan, L. 2014. Do male-female wage differentials reflect differences in the return to skill? Cross-city evidence from 1980–2000. American Economic Journal: Applied Economics 6 (2): 178–94.

Becker, Charles M. & Andrew R. Morrison. 1999. 'Chapter 43 Urbanization in Transforming Economies'. In , edited by BT - Handbook of Regional and Urban Economics, 3:1673–1790. Applied Urban Economics. Elsevier.

Bengoa, M., Sanchez-Robles, B., 2003. Foreign direct investment, economic freedom and growth: new evidence from Latin America. European Journal of Political Economy 19 (3): 529–545.

Bernanke, B. (1983). Nonmonetary effects of the financial crisis in propagation of the great depression. American Economic Review 79: 257–276.

Bernanke, B.S., Gertler, M. (1989). Agency costs, net worth, and business fluctuations. American Economic Review 79:14–31.

Bernard, A. B., Jensen, J. B., Redding, S. J. & Schott, P. K., 2011. The empirics of firm heterogeneity and international trade. Annual Review of Economics 4, 283–313.

Bertrand, M., 2011. Chapter 17 - New perspectives on gender. In Handbook of Labor Economics, edited by David Card & Orley Ashenfelter, 4, Part B: 1543–90. Elsevier.

Bezemer, D., Jong-A-Pin, R., 2013. Democracy, globalization and ethnic violence. Journal of Comparative Economics 41 (1): 108–125.

Black, J., de Meza, D., Jeffreys, D., 1996. House prices, the supply of collateral and the enterprise economy. Economic Journal 106 (343), 60–75.

Blau, F. D., 1998. Trends in the well-being of American women, 1970–1995. Journal of Economic Literature 36 (1): 112–165.

Blau, F. D., Lawrence M. K., 2006. The US gender pay gap in the 1990s: slowing convergence. Industrial and Labor Relations Review 60 (1): 45–66.

Blinder, A. S. 1973. Wage discrimination: Reduced form and structural estimates. The Journal of Human Resources 8 (4): 436–455.

Blundell, R., Bond, S., 1998. Initial conditions and moment restrictions in dynamic panel data models. Journal of Econometrics 87 (1): 115 – 143.

Bobba, M., Coviello, D., 2007. Weak instruments and weak identification in estimating the effects of education, on democracy. Economics Letters 96 (3): 301–306.

Borensztein, E., De Gregorio, J. & Lee, J-W., 1998. How does foreign direct investment affect economic growth? Journal of International Economics 45 (1): 115–135.

Brandt L, Van Biesebroeck J, Zhang Y. (2014) Challenges of working with the Chinese NBS firm-level data. China Economic Review, 30: 339-352.

Brandt, L., Biesebroeck, J & Zhang, Y. (2012). Creative accounting or creative destruction? Firm-level productivity growth in Chinese manufacturing. Journal of Development Economics, 97(2): 339–351.

Broxterman, D. A., Yezer, A. M., 2015. Why does skill intensity vary across cities? The role of housing cost. Regional Science and Urban Economics 55, 14–27.

Brueckner, J. K. 2011. Lectures on Urban Economics. Cambridge, Mass: MIT Press.

Buckles, K. S., Hungerman, D. M., 2012. Season of birth and later outcomes: Old questions, new answers. Review of Economics and Statistics 95 (3): 711–24.

Bulte, E., Xu, L. & Zhang, X., (2015). Does aid promote or hinder industrial development? Lessons from a natural experiment in China.

Cao, G., Feng, C., Tao, R., 2008. Local ”land finance” in China’s urban expansion: Challenges and solutions. China & World Economy 16 (2): 19–30.

Carmeron, C., Triveda, P., 2010. Microeconomics using Stata. Revised edition. Statacorp LP College Station, Texas.

Chan, S., Han, G. & Zhang, W., 2016. How strong are the linkages between real estate and other sectors in China? Research in International Business and Finance 36, 52–72.

Chaney, T., Sraer, D. & Thesmar, D., 2012. The collateral channel: how real estate shocks affect corporate investment. American Economic Review 102 (6): 2381–2409.

Chang, H., Dong, X-Y. & Macphail, F., 2011. Labor migration and time use patterns of the left-behind children and elderly in rural china. World Development 39(12): 2199-2210.

Charles, K. K., Guryan, J., 2011. Studying discrimination: fundamental challenges and recent progress. Annual Review of Economics 3 (1): 479–511.

Chen, K., Wen, Y., 2014. The great housing boom of China. FRB of St. Louis Working Paper No. 2014-22. Federal Reserve Bank, St. Louis.

Chen, P., Wang, C., Liu, Y., 2015. Real estate prices and firm borrowings: Micro evidence from China. China Economic Review 36, 296–308.

Chen, T., Liu, L. X. & Zhou, L-A., 2015. The crowding-out effects of real estate shocks–evidence from China. Available at SSRN 2584302.

Chen, T. & J.K.-S. Kung. 2016. Do Land Revenue Windfalls Create a Political Resource Curse? Evidence from China. Journal of Development Economics 123: 86–106.

Cole, M. A., Elliott, R. J. & Fredriksson, P. G., 2006. Endogenous pollution havens: Does FDI influence environmental regulations? Scandinavian Journal of Economics 108 (1): 157–178.

Corden, W. M. & Neary, J. P., 1982. Booming sector and de-industrialization in a small open economy. The Economic Journal, 92(368): 825-848.

Cui, E., Tao, R., Warner, T. J. & Yang, D., 2015. How do land takings affect political trust in rural China? Political Studies, 63 (S1): 91-109.

Cvijanovic, D., 2014. Real estate prices and firm capital structure. Review of Financial Studies 27, 2690–2735.

Dang, D. A., 2013. How foreign direct investment promote institutional quality: Evidence from Vietnam. Journal of Comparative Economics 41 (4): 1054–1072.

Dean, J. M., Lovely, M. E., Wang, H., 2009. Are foreign investors attracted to weak environmental regulations? Evaluating the evidence from China. Journal of Development Economics 90 (1): 1–13.

Deaton, A., 1999. Commodity prices and growth in Africa. Journal of Economic Perspectives 13 (3): 23–40.

Démurger, S., Gurgand, M., Li, S. & Yue, X., 2009. Migrants as second-class workers in urban China? A decomposition analysis. Journal of Comparative Economics 37 (4): 610–28.

Deng, Y., Gyourko, J. & Wu, J., 2012. Land and house price measurement in China. In: Heath, A., Packer, F., Windsor, C. (Eds.), Property Markets and Financial Stability. Reserve Bank of Australia, Sydney.

Dickens, W., Katz L., 1987. Inter-industry wage differences and industry characteristics, in K. Lang & J. Leonard (Eds.), Unemployment and the Structure of Labor Markets. Oxford, Basil Blackwell, pp. 28-89.

Duranton, Gilles & Diego Puga. 2004. Chapter 48 - Micro-Foundations of Urban Agglomeration Economies. In Handbook of Regional and Urban Economics, edited by J. Vernon Henderson and Jacques-François Thisse, 4:2063–2117. Cities and Geography. Elsevier.

Durlauf, S. N., Johnson, P. A. & Temple, J. R., 2005. Chapter 8 - Growth econometrics. In: Aghion, P., Durlauf, S. N. (Eds.) Handbook of Economic Growth, Volume 1: 555–677. Amsterdam: Elsevier.

Edmond, C., Midrigan, V. & Xu, D. Y., 2015. Competition, markups, and the gains from international trade. American Economic Review 105, 3183–3221.

Ellison, G. & Glaeser, E. L., 1997. Geographic concentration in U.S. manufacturing industries: A dartboard approach, Journal of Political Economy 105, 889-927.

Feenstra, R. C., Li Z. & Yu M., 2014. Exports and credit constraints under incomplete information: theory and evidence from China. Review of Economics and Statistics 96 (3): 729–744.

Feltenstein, A., Iwata, S., 2005. Decentralization and macroeconomic performance in China: regional autonomy has its costs. Journal of Development Economics 76 (2): 481 – 501.

Fernández, J E. 2007. Resource Consumption of New Urban Construction in China. Journal of Industrial Ecology 11 (2): 99–115.

Fortin, N. M. 2005. Gender role attitudes and the labor-market outcomes of women across OECD countries. Oxford Review of Economic Policy 21 (3): 416–38.

Fortin, N. M., Lemieux, T. & Firpo, S. 2011. Chapter 1 - Decomposition methods in economics. In: Handbook of Labor Economics, edited by Orley Ashenfelter and David Card, 4, Part A:1–102. Elsevier.

Gan, J., 2007a. Collateral, debt capacity, and corporate investment: evidence from a natural experiment, Journal of Financial Economics 85: 709-734.

Gan, J., 2007b. The real effects of asset market bubbles: loan- and firm-level evidence of a lending channel, Review of Financial Studies 20: 1941-1973.

Gardeazabal, J., Ugidos, A., 2004. More on identification in detailed wage decompositions. The Review of Economics and Statistics 86: 1034–1036.

Gelbach, J., 2014. B1X2: Stata module to account for changes when X2 is added to a base model with X1. Statistical Software Components. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2615909.

Glaeser, E. L. 1999. Learning in Cities. Journal of Urban Economics 46 (2): 254–77.

Glaeser, E. L. & David C. Maré. 2001. Cities and Skills. Journal of Labor Economics 19 (2): 316–42.

Glaeser, E. L. & Kahn, M. E., 2004. 'Chapter 56 - Sprawl and Urban Growth'. In Handbook of Regional and Urban Economics, edited by J. Vernon Henderson and Jacques-François Thisse 4:2481–2527. Cities and Geography. Elsevier.

Glaeser, E., Huang, W., Ma, Y. & Shleifer, A., 2016. A Real Estate Boom with Chinese Characteristics. Journal of Economic Perspectives.

Glaeser, E., 2011. Triumph of the City: How Our Greatest Invention Makes Us Richer, Smarter, Greener, Healthier, and Happier. Penguin Books.

Goldberg, P. K., Pavcnik, N., 2007. Distributional effects of globalization in developing countries. Journal of Economic Literature 45: 39–82.

Grove, W. A., Hussey, A., & Jetter, M., 2011. The gender pay gap beyond human capital heterogeneity in noncognitive skills and in labor market tastes. Journal of Human Resources 46 (4): 827–74.

Guryan, J., Charles, K. K., 2013. Taste-based or statistical discrimination: the economics of discrimination returns to its roots. The Economic Journal 123 (572): F417–32.

Gwartney, J., Lawson, R., Hall, J., 2013. Economic freedom of the world: 2013 annual report. Fraser Institute, Vancouver. http://www.freetheworld.com/.

Hayakawa, K., 2007. Small sample bias properties of the system GMM estimator in dynamic panel data models. Economics Letters 95 (1): 32-38

Heath, R. & Mobarak, A. M., 2015. Manufacturing growth and the lives of Bangladeshi women. Journal of Development Economics 115 (July): 1–15.

Hsieh, C. & Klenow, P. 2009. Misallocation and manufacturing TFP in China and India. The Quarterly Journal of Economics 124 (4): 1403-1448. URL: *http://media.wix.com/ugd/02822d_356110380935490f9888895d4d601361.pdf*

Hsieh, Chang-Tai & Enrico Moretti. 2015. Why Do Cities Matter? Local Growth and Aggregate Growth. Working Paper 21154. National Bureau of Economic Research.

Jefferson, G. H, Rawski, T. & Zhang, Y., 2008. Productivity growth and convergence across China's industrial economy. Journal of Chinese Economics and Business Studies 6(2): 121-140.

Jin, H., Qian, Y. & Weingast, B. R., 2005. Regional decentralization and fiscal incentives: federalism, Chinese style. Journal of Public Economics 89 (9–10): 1719–1742.

Juhn, C., Ujhelyi, G. & Villegas-Sanchez, C., 2013. Trade liberalization and gender inequality. American Economic Review 103 (3): 269–73.

Juhn, C., Ujhelyi, G. & Villegas-Sanchez, C., 2014. Men, women, and machines: How trade impacts gender inequality. Journal of Development Economics 106: 179–93.

Kiyotaki, N. & Moore, J., 1997. Credit chains. Journal of Political Economy, 105(21): 211-248.

Kong, S. T., 2010. Rural-urban migration in China: Survey design and implementation. In: Meng, Xin & Manning, Chris (Eds.) with Shi, Li & Effendi, Tadjuddin, The Great Migration: Rural-Urban Migration in China and Indonesia, Edward Elgar.

Kong, S. T., Meng X., Zhang D., 2009. Impact of economic slowdown on migrant workers. In: Garnaut R, Song L, Woo W-T (Eds) China's New Place in a World in Crisis:

Kotsadam, A. & Tolonen, A., 2016. African mining, gender, and local employment. World Development 83: 325–339.

Kung, J. K. & Chen, T., 2014. Do land revenue windfalls create a political resource curse? Evidence from China. URL: http://media.wix.com/ugd/02822d_356110380935490f9888895d4d601361.pdf.

Laeven, L., Woodruff, C., 2007. The quality of the legal system, firm ownership, and firm size. Review of Economics and Statistics 89: 601–614.

Lan, J., Kakinaka, M., Huang, X., 2012. Foreign direct investment, human capital and environmental pollution in China. Environmental and Resource Economics 51 (2): 255–275.

Lemieux, Thomas. 2006a. Increasing Residual Wage Inequality: Composition Effects, Noisy Data, or Rising Demand for Skill? American Economic Review 96 (3): 461–98.

Lemieux, Thomas. 2006b. The "Mincer Equation" Thirty Years After Schooling, Experience, and Earnings. In Jacob Mincer A Pioneer of Modern Labor Economics, edited by Shoshana Grossbard, 127–145. Boston, MA: Springer US.

Levinsohn & J., Petrin, A., 2003. Estimating production functions using inputs to control for unobservables. Review of Economic Studies 70: 317-341.

Li, L. & Wu, X., 2014. Housing price and entrepreneurship in China. Journal of Comparative Economics, 42 (2): 436-449.

Lin, J. Y. & Liu, Z., 2000. Fiscal decentralization and economic growth in China. Economic Development and Cultural Change 49 (1): 1–21.

Liu, M., Tao, R., Yuan, F. & Cao, G., 2008. Instrumental land use investment-driven growth in China. Journal of the Asia Pacific Economy 13 (3): 313–331.

Liu, Q. & Lu, Y., 2015. Firm Investment and Exporting: Evidence from China's Value-Added Tax Reform. Journal of International Economics 97 (2): 392–403.

Liu, Z., Wang, P., Zha, T., 2013. Land-price dynamics and macroeconomic fluctuations. Econometrica 81, 1147–1184.

Lu, Y., Tao Z. & Yu, L.,2012. Agglomeration and Markup. Available at SSRN 2065303.

Lu, X., Li, P. & Chen, G. (Eds.), 2012. Society of China: analysis and forecast (2013). Blue Book of China's Society. Beijing: Social Science Academic Press. [In Chinese].

Lu, Y. & Yu, L., 2015. Trade liberalization and markup dispersion: evidence from China's WTO accession. American Economic Journal: Applied Economics 7, 221-253.

Magnani, E. & Zhu, R., 2012. Gender wage differentials among rural–urban migrants in China. Regional Science and Urban Economics, Special issue on Migration and Development, 42 (5): 779–93.

Magnani, E. & Zhu, R., 2015. Wage Differentials between urban residents and rural migrants in urban China during 2002–2007: A distributional analysis. China Economic Review.

Majlesi, K., 2016. Labor market opportunities and women's decision making power within households. Journal of Development Economics 119: 34-47.

Manning, A. & Swaffield, J., 2008. The gender gap in early-career wage growth. The Economic Journal 118 (530): 983–1024.

Martin, P., Mayer, T. & Thoenig, M., 2008. Make trade not war? Review of Economic Studies 75 (3): 865–900.

Martinez-Vazquez, J. & McNab, R. M., 2003. Fiscal decentralization and economic growth. World Development 31 (9): 1597 – 1616.

Mehlum, H., Moene, K. & Torvik, R., 2006. Institutions and the resource curse. Economic Journal 116 (508): 1–20.

Meligrana, J., Li, Z. & Zhang, Z., 2011. Resolving land use disputes in China: An analysis of a method of dealing with citizen complaints. Environment and Urbanization Asia 2 (2): 251–264.

Melitz, M. J., 2003. The impact of trade on intra-industry reallocations and aggregate industry productivity. Econometrica 71: 1695-1725.

Meng, X., 2012. Labor market outcomes and reforms in China. Journal of Economic Perspectives 26 (4): 75–102.

Meng, X. & Zhang, J., 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.

Meng, X. & Miller, P., 1995. Occupational segregation and its impact on gender wage discrimination in China's rural industrial sector. Oxford Economic Papers 47 (1): 136–55.

Mian, A. & Sufi, A., 2011. House prices, home equity–based borrowing, and the US household leverage crisis. American Economic Review 101: 2132–56.

Michaels, G. (2011). The long term consequences of resource-based specialisation. The Economic Journal, 121(551): 31-57.

Miguel, E., Satyanath, S. & Sergenti, E. (2004). Economic shocks and civil conflict: An instrumental variables approach. Journal of Political Economy 112(4): 725-753.

Ministry of Land and Resources of China (MLR) (2000 – 2011). China land and resources almanac (2000-2011). Beijing: Geological Publishing House.

National Bureau of Statistics (NBS) (2011). Demographic statistics from the 2010 Census. National Bureau of Statistics of China. [In Chinese]

National Bureau of Statistics of China (NBSa) (2000 – 2011). China Statistical Yearbook (2000 - 2011). Beijing: China Statistics Press.

National Bureau of Statistics of China (NBSb) (2000 – 2011). China Statistical Yearbook for Regional Economy (2000 - 2011). Beijing: China Statistics Press.

Neumark, D., 1988. Employers' discriminatory behavior and the estimation of wage discrimination. The Journal of Human Resources 23 (3): 279–95.

Nillesen, E. & Bulte, E., 2014. Natural resources and violent conflict. Annual Review of Resource Economics 6 (1): 69–83.

Nitikin, D., Shen, C. & Zou, H.-f., 2012. Land taxation in China: Assessment of prospects for politically and economically sustainable reform. Annals of Economics and Finance 13 (2): 489–528.

Nordman, C. J. & Wolff, F.C., 2009. Is there a glass ceiling in morocco? Evidence from matched worker–firm data. Journal of African Economies

Nordman, C. J., Robilliard, A. S. & Roubaud, F., 2011. Gender and ethnic earnings gaps in seven West African cities. Labor Economics, Labor Markets in Developing Countries, 18: S132–45.

O'Sullivan, A., 2012. Urban Economics. 8th ed. New York, NY: McGraw-Hill / Irwin.

Oaxaca, Ronald L & Michael R Ransom. 1994. On discrimination and the decomposition of wage differentials. Journal of Econometrics 61 (1): 5–21.

Oaxaca, Ronald. 1973. Male-female wage differentials in urban labor markets. International Economic Review 14 (3): 693–709.

Pils, E., 2005. Land disputes, rights assertion, and social unrest in China: a case from Sichuan. Columbia Journal of Asian Law 19, 235.

Qiao, B., Martinez-Vazquez, J. & Xu, Y., 2008. The tradeoff between growth and equity in decentralization policy: China's experience. Journal of Development Economics 86 (1): 112 – 128.

Qin, Min, James J. Brown & Sabu S. Padmadas, Bohua Li, Jianan Qi, Jane Falkingham. 2016. Gender inequalities in employment and wage-earning among internal labor migrants in Chinese cities. Demographic Research 34: 175-202

Restuccia, D. & Rogerson, R., 2008. Policy distortions and aggregate productivity with heterogeneous establishments. Review of Economic Dynamics 11: 707-720.

Robertson, G. B. & Teitelbaum, E., 2011. Foreign direct investment, regime type, and labor protest in developing countries. American Journal of Political Science 55 (3): 665–677.

Robinson, J. A., Torvik, R. & Verdier, T., 2006. Political foundations of the resource curse. Journal of Development Economics 79 (2): 447–468.

Rong, Z., Wang, W. & Gong, Q., 2016. Housing price appreciation, investment opportunity, and firm innovation: evidence from China. Journal of Housing Economics 33: 34–58.

Roodman, D., 2009. How to do xtabond2: An introduction to difference and system GMM in Stata. Stata Journal 9 (1): 86 -136.

Ross, M. L., 2015. What have we learned about the resource curse? Annual Review of Political Science, 18: 239-259.

Ryzhenkov, M., 2016. Resource misallocation and manufacturing productivity: the case of Ukraine. Journal of Comparative Economics 44: 41–55.

Saiz, A. 2010. The geographic determinants of housing supply. The Quarterly Journal of Economics, 125(3): 1253-1296.

Schmalz, M. C., Sraer, D. A. & Thesmar, D., 2015. Housing collateral and entrepreneurship. Journal of Finance, forthcoming.

Scholz, J. K. & Sicinski, K., 2015. Facial attractiveness and lifetime earnings: Evidence from a cohort study. Review of Economics and Statistics, 97(1): 14-28.

Shiller, Robert J., 2015. Irrational Exuberance. Princeton University Press, Princeton.

Song, Z. & Wu. G., 2012. A structural estimation on capital market distortions in Chinese manufacturing, mimeo, University of Chicago.

Swanson, A. 2015. How China used more cement in 3 years than the U.S. did in the entire 20th Century. *https://www.washingtonpost.com/news/wonk/wp/2015/03/24/how-china-used-more-cement-in-3-years-than-the-u-s-did-in-the-entire-20th-century/?utm_term=.9112b6f72d6e*

(March 24, 2015)

Syverson, Chad, 2004. Product substitutability and productivity dispersion. Review of Economics and Statistics 86, 534-550.

Syverson, Chad, 2011. What determines productivity? Journal of Economic Literature 49, 326–365.

Tan, R., Qu, F., Heerink, N. & Mettepenningen, E., 2011. Rural to urban land conversion in China — How large is the over-conversion and what are its welfare implications? China Economic Review 22 (4): 474–84.

Tao, R., Su, F., Liu, M. & Cao, G., 2010. Land leasing and local public finance in China's regional development: Evidence from prefecture-level cities. Urban Studies 47 (10): 2217–2236.

The Economist (2014) Building the dream *http://www.economist.com/news/special-report/21600797-2030-chinese-cities-will-be-home-about-1-billion-people-getting-urban-china-work*

The Economist (2015) The paradox of soil http://www.economist.com/news/briefing/21647622-land-centre-pre-industrial-economy-has-returned-constraint-growth

Thornton, J., 2007. Fiscal decentralization and economic growth reconsidered. Journal of Urban Economics 61 (1): 64 – 70.

UN HABITAT. 2016. World Cities Report 2016: Urbanization and development. Emerging futures. Nairobi, Kenya: UN HABITAT.

UNCTAD, 2012. World Investment Report. United Nations Press, New York and Geneva.

Van der Ploeg, Rick & Anthony J. Venables. Natural resource wealth: the challenge of managing a windfall. Annual Review of Economics 4 (2012): 315-337.

Wang, S.-Y., 2012. Credit constraints, job mobility, and entrepreneurship: evidence from a property reform in China. Review of Economics and Statistics 94, 532–551.

Wang, Y., 2013. Fiscal decentralization, endogenous policies, and foreign direct investment: Theory and evidence from China and India. Journal of Development Economics 103, 107 – 123.

Wang, Ya Ping, Murie, Alan, 1996. The process of commercialization of urban housing in China. Urban Studies 33, 971–89.

Wooldridge, J., 2012. Introductory econometrics: A modern approach, 5th Edition. South-Western College Publishing, Cincinnati, OH.

Wu, Guiying Laura, Feng, Qu, Li, Pei, 2015. Does local governments' budget deficit push up housing prices in China? China Economic Review 35, 183–196.

Wu, J., Deng, Y., Liu, H., 2014. House price index construction in the nascent housing market: the case of China. Journal of Real Estate Finance and Economics 48: 522–545.

Wu, J., Gyourko J & Deng, Y., 2015. Real estate collateral value and investment: The case of China. Journal of Urban Economics 86: 43-53.

Wu, J., Gyourko, J. & Deng, Y., 2012. Evaluating conditions in major Chinese housing markets. Regional Science and Urban Economics 42: 531–43.

Wu, Jing, Joseph Gyourko & Yongheng Deng. 2012. Evaluating Conditions in Major Chinese Housing Markets. Regional Science and Urban Economics, Special Section on Asian Real Estate Market, 42 (3): 531–43.

Xu, C., 2011. The fundamental institutions of China's reforms and development. Journal of Economic Literature 49 (4): 1076–1151.

Yang, D. Y.-R. & Wang, H.-K., 2008. Dilemmas of local governance under the development zone fever in China: A case study of the Suzhou region. Urban Studies 45 (5-6): 1037–1054.

Zhang, X., 2006. Fiscal decentralization and political centralization in China: Implications for growth and inequality. Journal of Comparative Economics 34 (4): 713 – 726.

Zheng, S. & Matthew E. Kahn. China's bullet trains facilitate market integration and mitigate the cost of megacity growth, 2013. Proceedings of the National Academy of Sciences 110.14: E1248-E1253.

Zheng, X., Li, F., Song, S. & Yu, Y., 2013. Central government's infrastructure investment across Chinese regions: A dynamic spatial panel data approach. China Economic Review 27: 264–276.

Zhu, R. 2015. Heterogeneity in the economic returns to schooling among Chinese rural–urban migrants, 2002–07. Economics of Transition 23 (1): 135–67.

Zhu, Rong. 2015. Wage Differentials between Urban Residents and Rural Migrants in Urban China during 2002–2007: A Distributional Analysis. China Economic Review.

Summary

In this dissertation I address emerging land and labor issues associated with the rapid urbanization in China over the past decades, including the negative effects of the real estate booms on manufacturing development, land conflicts, and gender inequality among migrants. To understand the causes of the problems, I use a range of data sets at firm, industrial, regional, and individual/household levels to investigate the potential causes of the emerging challenges associated with urbanization.

The thesis consists of six chapters. Chapter 1 offers an introductory discussion on the overarching objective and specific questions of the whole research. Chapters 2—5 are four (self-contained) empirical papers — each addressing one research question. They can be divided into two parts. Part 1 consists of Chapters 2 and 3, which focus on the impacts of the real estate boom accompanying rapid urbanization in China on the industrial economy. Part 2 consist of Chapters 4 and 5, which address two socio-economic challenges — land conflicts and gender inequality, respectively. Chapter 6 concludes the thesis with discussions of the main findings and recommendations for policies and future research.

Chapter 2 focuses on the effects of housing price appreciation on firm investment. Using a comprehensive dataset of all medium and large enterprises in China between 2003 and 2007, I find a robust negative relationship between local housing prices and the level of investment of manufacturing firms. A detailed examination of the underlying mechanism probes that it is mainly due to a Dutch disease effect of the real estate price boom: rising housing prices push up wages and other production costs for manufacturing firms and therefore reduce the incentive to invest.

Chapter 3 applies a difference-in-differences approach to identify the intra-industry resource allocation effects of the real estate boom. The results show that industries intrinsically highly linked with real estate sectors experienced increasing heterogeneity in firm total factor productivity (TFP), suggesting sorting-in or expansion of less-efficient firms in more-exposed industries.

Chapter 4 investigates the surging of land conflicts in the recent decades. The chapter uses (provincial) data on illegal land uses during the period 1999–2010 as a proxy for the intensity of land conflicts to investigate the effects of foreign direct investment and fiscal decentralization on jurisdictional land conflicts. The results show that the FDI growth rate has a positive and significant impact on the growth rate of illegal land use when there is a high degree of fiscal decentralization. It provides fresh evidence that FDI inflows trigger tensions over land in provinces with a high degree of fiscal decentralization.

Chapter 5 examines the wage gap between female and male rural-urban migrant workers in China. It tests whether the differences can be attributed to human capital characteristics, gender discrimination or other factors. I document a relative small gender wage gap of 16-18%, and find that most of the gender wage gap cannot be attributed to gender differences in observed characteristics. In addition, the paper finds important differences in factors affecting gender wage gaps between the sub-sample of migrants living at their workplace and those living at urban communities.

Chapter 6 concludes. The main findings in this thesis indicate that the problems China's urbanization met are largely due to different kinds of market and policy failures. Accordingly, the policies need deal with these failures, in particular, addressing the negative externalities of the rising housing prices on industrial development, improving the institution environments of urbanization, and investing in the equal access to urban labor market.

Acknowledgements

My PhD journey, most of the time, was enjoyable. Learning new theories, reading top publications, and acquiring new data analysis skills were particularly entertaining. There were also a lot of moments I was stuck with the lack of ideas and the shortage of data. I thank my promoter Prof. Erwin Bulte and supervisor Dr. Nico Heerink for their guidance and support, especially at those difficult moments. I remember that the first time I discussed with Erwin on my research proposal. Erwin printed a few papers and put them on my desk after our meeting; they were on theories of jurisdictional tax competition. Since then Erwin have been helping me a lot in getting my theories right. I often feel sorry for Nico as he needed to correct all my writing from the title to the last reference of every paper I wrote. Nico also invested a lot of time and patience to deal with my unstable ideas and writings. I doubt if I can pay the academic debts back. My thanks also go to Prof. Shi Xiaoping and Prof. Qu Futian; without their support and training when I was in Nanjing, I may have no chance to open a door of academic research. If I did learn something in the past years, it is from the discussions and communications with them.

I was very fortunate to meet Prof. Tu Qin, Prof. Fu Shihe, Prof. Tao Ran, Prof. Feng Shuyi, Dr. Janneke Pieters, Dr. Yu Linhui, Dr. Mao Risheng, and Dr. Wang Jianghao. They have helped me shape the research in this thesis or directly contributed to the writing of the papers included in this thesis. My study and research also benefit from many colleagues at Development Economics Group—at the coffee table or in the corridor. They are of diverse (academic) backgrounds, and sometimes are from different continents. The diversity of the DEC group enables a multi-perspective thoughts on economics (and life) — a treasure that few other places have.

Also, many other people have helped me in many ways. To name but a few, the secretaries, the football mates, the basketball mates on Thursday evenings, and the Chinese friends cooking together on weekends are unforgettable.

Last but not least, I thank Mingyan and Dingding, for their love, support and endurance. They gave me many understandings when I could not give what they deserved.

Land and Labor in Urbanizing China

Yan Wu
Wageningen School of Social Sciences (WASS)
Completed Training and Supervision Plan

Wageningen School of Social Sciences

Name of the learning activity	Department/Institute	Year	ECTS*
A) Project related competences			
Advanced Macroeconomics, ENR 30806	WUR	2009	6
Dynamic Macroeconomics Theory	NAKE	2009	3
Panel Data Analysis in Microeconomics	WASS	2009	4
Rural Economic Analysis, AEP 31306	WUR	2010	6
Spatial Econometrics	WASS	2011	1.5
B) General research related competences			
Mansholt Introduction Course	WASS	2009	1.5
Research Methodology: from Topic to Proposal	WASS	2010	4
Techniques for Writing and Presenting a Scientific Paper	WGS	2009	1.2
C) Career related competences/personal development			
'FDI, fiscal decentralization and land conflicts in China'	9th International Conference on Chinese Economy, University of Auvergne, France	2013	1
'Does the real estate boom hinder industrial development? Evidence from Chinese Cities'	3rd RUSE Workshop, Fudan University, China	2014	1
'Local effects of housing price appreciation on industrial investment: firm-level evidence from China'	4th RUSE Workshop, Tsinghua University, China	2015	1
'The gender wage gap in China's rural-urban migrants' labor market: New evidence'	8th International Symposium on Human Capital and Labor Markets, Beijing, China	2016	1
D) Other advanced courses (Qualifying examination courses attained at WUR)			
Advanced Microeconomics, ECH 32306	WUR	2009	-
Advanced Econometrics, AEP 60306	WUR	2009	-
Total			**31.2**

*One credit according to ECTS is on average equivalent to 28 hours of study load.

Funding

This research described in this thesis was financially supported by the Programme Strategic Scientific Alliances (PSA) of the Royal Netherlands Academy of Arts and Sciences (KNAW) and the Ministry of Science and Technology of PR China (08 -PSA-E-02).

Financial support from the Wageningen School of Social Science and the Erasmus Mundus Programme (the Eurasia European Academic Mobility Network with Asia 141210-EM-1-2008-1-AT-ERAMUNDUS-ECW-L14) of the European Commission is gratefully acknowledged.