UNCOVERING THE MECHANISM UNDERLYING THE EDUCATIONAL SELF-SELECTION OF INTERNAL MIGRANTS IN CHINA

ZHOU JIANTAO

PhD

The Hong Kong Polytechnic University

2022

The Hong Kong Polytechnic University

Department of Building and Real Estate

Uncovering the Mechanism Underlying the Educational Self-Selection of Internal Migrants in China

ZHOU Jiantao

A thesis submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

March 2022

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

_______________(Signed)

ZHOU Jiantao (Name of student)

ABSTRACT

In today's knowledge-based economy, human resources play an increasingly critical role in regional development compared to the cheap labor of the past. Consequently, as an essential source of labor supply, the role of population migration in economic growth is also changing. The agglomeration of high-skilled labor facilitates the economic performance of destination regions. In turn, the loss of this labor undermines the economy of original regions, which enlarges the regional development gap. In such cases, instead of the scale of migration, it is necessary to obtain more understanding about the skill composition of migrants and its underlying determining mechanism, while classical "push and pull" theories fail to explain it. This problem is particularly significant in China, given that China has undergone radical changes in recent decades, from a labor-intensive economy in the beginning, that relied on large amounts of cheap rural labor, to the industrial upgrading reform in the last decade, that has created an enormous demand for highly skilled labor.

The prerequisite for understanding the skill composition of migrants is to realize the self-selection mechanism of migrants, which is hardly discussed in China. This thesis aims to fill this research gap and comprehensively investigates the self-selection of migrants in China and its underlying mechanisms. To achieve this goal, this research first empirically portrays the self-selection pattern of migrants, employing four periods of Census data. Then, the classical self-selection framework is extended to explain the underlying mechanism by introducing four new ingredients: the household registration (*hukou*) system, inequality of opportunity,

technological change, and housing costs. Finally, this research proposes several empirical econometric models to verify proposed four new impact channels that induce migrant self-selection. The major findings are discussed below.

The most important finding of this research is that the internal migration in China shows a U-shaped selection pattern, unlike most cases found in other regions/countries. High- and low-skilled individuals have higher propensities to migrate (captured by migration rates), while mid-skilled ones surprisingly have the lowest. A key explanatory factor is China's unique *hukou* system. This research theoretically and empirically verifies that the *hukou* system reduces migrants' income levels through labor market discrimination on the one hand, and increases migrants' living costs by limiting social benefits on the other. As a result, given the skill-biased local *hukou* application mechanism (which prefers high-skilled migrants), this system has reshaped the migrant selection pattern and led to a U-shaped one.

In addition to the *hukou* system, the difference in regional return to skills and heterogeneous migration costs also contribute to this pattern. First, this research finds that developed coastal regions have relatively lower income inequality (representing the return to skills) and thus asymmetrically attract more low-skilled migrants than high-skilled ones. Besides, two income inequality components, inequality of opportunity (induced by uneven social opportunities) and inequality of effort (induced by varying personal efforts), also lead to a positive selection of migrants. This result implies that those low-skilled migrants from inland to coastal regions are not only chasing economic returns but also more social opportunities.

Second, however, the attractiveness of developed coastal regions for low-skilled migrants is fading due to the labor market shock induced by technological change. This research takes the industrial robot installation as a case to investigate how technological change alters the labor market and influences the skill demand for migrant labor. The results show that cities with higher levels of robot exposure attract more high-skilled migrants in production sectors but crowd out low-skilled ones, which implies that industrial robots mainly displace low-skilled labor but need more complements from high-skilled labor. This mechanism has resulted in a significant positive selection of migrants.

Thirdly, migrants are also facing housing unaffordability issues that significantly increase their migration costs, along with the economic development in destination cities. This research theoretically and empirically explores the heterogeneous effects of housing costs on migrants with varying skill levels. The results show that low-skilled migrants are crowded out from big cities due to relatively higher housing costs, similar to technological change. As a result, except for the *hukou* system and technological change, the housing prices also build a barrier to select high-skilled migrants settling in developed regions but crowd out low-skilled ones to other less-developed regions.

LIST OF PUBLICATIONS

➢ Peer-reviewed papers arising from the thesis

Zhou, J.*, & Hui, E. C. M. (2022). The *hukou* system and selective internal migration in China. *Papers in Regional Science*, *101*(2), 461-482.

Zhou, J.*, Hui, E. C. M., & Peng, H. (2022). Chasing opportunity? Inequality of opportunity and educational self-selection of interprovincial migrants in China. *The Annals of Regional Science*, 1-29.

Zhou, J.*, & Hui, E. C. M. (2022). Housing prices, migration, and self-selection of migrants in China. *Habitat International*, 119, 102479.

➢ Papers under peer review arising from the thesis

Zhou, J.*, & Hui, E. C. M. "How do *hukou* restrictions affect the skill composition of urban migrants in China?" (Revising)

Zhou, J.*, Hui, E. C. M., & Peng, H. "Robots and skill sorting in urban China: which migrant workers are crowded out by industrial robots?" (Revising)

➢ Other publications

Zhang, L., **Zhou, J.***, & Hui, E. C. M. (2020). Which types of shopping malls affect housing prices? From the perspective of spatial accessibility. *Habitat International*, 96, 102118.

Zhang, L., **Zhou, J.**, Hui, E. C., & Wen, H. (2019). The effects of a shopping mall on housing prices: A case study in Hangzhou. *International Journal of Strategic Property Management*, 23(1), 65-80.

ACKNOWLEDGEMENTS

I sincerely thank the Department of Building and Real Estate, the Hong Kong Polytechnic University, for giving me this valuable opportunity to pursue my Ph.D., without which I would not have gained many valuable research experiences, academic mentors, and close friends.

I would like to express my gratitude to many people during this three-year journey. The person I want to thank most is my chief supervisor, Prof. Eddie Chi-Man Hui. He helped me open the door to the academia with his constant, selfless help, and guidance. It was his encouragement and advice that gave me the courage to take on this unfamiliar research topic. He likewise teaches me how to conduct good research, including finding a sound research gap, applying suitable research methods, and constructing a smooth writing expression.

I would also like to thank many friends in Hong Kong and mainland China, including Guangpu Xu, Tingting Ji, Huiwen Peng, Shuyuan Li, Zhenhua Huang, Zhenyu Zhang, Zhongnan Ye, and Yi Jian. My three years in Hong Kong have not been smooth, as several far-reaching events happened that have impacted my life and academic research. Because of these precious friends, I had the moral support to get through those tough times.

Last but not least, I would like to thank my parents and my brother. They have given me infinite love and encouragement and have been strong support in my academic and life path.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1. INTRODUCTION

1.1. Research background

1.1.1. China’s great migration in the past few decades

Since the reform and opening up in 1978, China has experienced rapid economic development and witnessed the world’s unprecedented ‘Great Migration,’ whereby an estimated 200-250 million rural residents moved to cities and towns within China (Chan, 2012). Such a large scale of migration has played a pivotal role in China's labor-intensive, export-led economic growth, especially after entering the World Trade Organization in 2001. Large amounts of idle labor from China's rural and inland areas are being drawn to coastal urban areas to fill the huge demand for cheap labor in labor-intensive manufacturing and supporting service industries. As shown in Figure 1.1, in parallel with the rapid growth of GDP, the size of China's floating population is also expanding rapidly. Freeman (2015) shows that 35% of China’s total workforce in 2015 was accounted for by migrants, a higher level than the share of migrants in the entire Chinese population. As a result, this labor redistribution contributed significantly to the urbanization process (Zhang & Song, 2003). According to data from the *China Statistical Yearbook 2019*, in the last four decades, China has experienced rapid growth in urbanization, from 17.58% in 1982 to 60.6% in 2019.

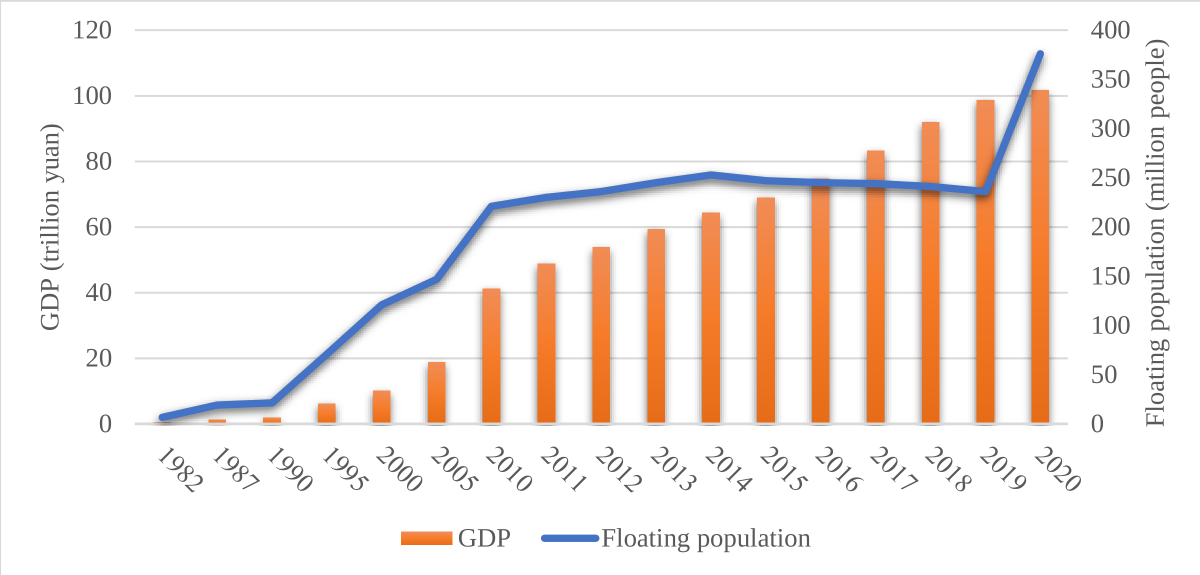

Figure 1.1 GDP and floating population growth in China from 1982 to 2020

Notes: 1. the data comes from the *China Statistical Yearbook 2021*; 2. the "floating population" is defined as people who have been away from their place of household registration (*hukou*) for more than six months, according to the China Bureau of Statistics (CBS).

However, the heavy reliance on cheap labor is gradually waning as a result of the ongoing industrial upgrading that China is undertaking. Central government continues to roll out plans and targets, including the "innovation-driven economy" in 2012 and "Made in China 2025" in 2015, which calls for the economic transition from labor-intensive to knowledge-intensive. In response to this call, local governments employ various means to encourage enterprises to upgrade their industries, which results in a dramatic increase in digitalization and automation across industries. Taking the industrial robots as an example, in 2016, China became the world's largest user of industrial robots, with nearly 350,000 units of industrial robots in use.

In such a case, what is more important for the regional economy is not the size of the migration, but the skill level of the migrants, given the increasingly important role of human capital in fostering economic transformation and development globally. On the one hand, the agglomeration of human capital in developed regions can directly promote regional economic performance, knowledge spillover, and innovation (Glaeser & Maré, 2001; Glaeser & Resseger, 2010; Lucas, 1988). On the other hand, “brain gain” in developed regions may in turn weaken regional convergence as it undermines the development of original less-developed regions (Fratesi & Percoco, 2014; Ganong & Shoag, 2017; Kanbur & Rapoport, 2005; Østbye & Westerlund, 2007). To attract more talents and enhance human capital accumulation, enterprises, cities, and even countries all over the world have successively issued ‘attraction policies’ to compete for talent and thus accelerate local economies (de Haas et al., 2016; de Lange et al., 2021; Haddad, 2020; Koslowski, 2014; Yang & Pan, 2020).

Along with the great strides of economic development, many social problems have also emerged in China, such as sky-rocketing housing prices and urban environmental issues, which profoundly influence the cost of migration (Chen et al., 2022; Zang et al., 2015). From 2002 to 2018, the average housing prices of four 1-tier cities in China have risen 8.4 (*Beijing*), 7.2 (*Shanghai*), 5.4 (*Guangzhou*), and 10.5 (*Shenzhen*) times. Given the worse risk tolerance, low-skilled and low-income migrants are suffering disproportionately from these social issues.

In the meantime, the unique household registration (*hukou*) system[1] is still deeply embedded

[1] The China’s *hukou* system is a household-based population management system which divides all population into agricultural and non-agricultural categories since 1958. At first, it strictly restricted the free movement of population between locations. After 1980s, the free movement of population is loosened, but social amenities are still deeply tied with their *hukou* locations. In 2014, the promulgation of “Opinions on further promoting the reform of the household registration system” further promoted the reform of this system (see official document,

in people's lives and affects their living costs, social welfare, and migration decisions (Bao et al., 2011; Fu & Ren, 2010; Song, 2014), although reform has been ongoing since the 1980s. As shown in Figure 1.2, the massive gap between the *de facto* and *hukou* urbanization rate still exists, implying that there are still many (low-skilled/low-income) migrants who cannot get the security of local *hukou*.

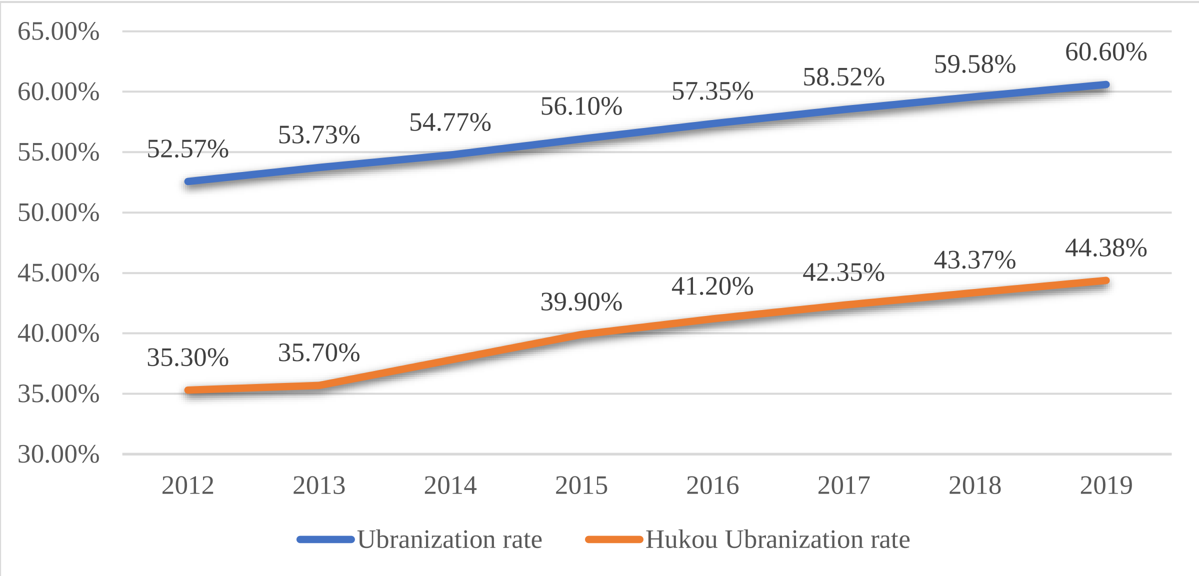

Figure 1.2 Urbanization rate, 2012-2019

Note: The data comes from *China Statistical Yearbook 2020*.

Under these social changes, the general pattern analysis of all migrants, that is the mainstream of previous studies on population migration (Liu & Shen, 2014; Molloy et al., 2011; Shen & Liu, 2016; Sjaastad, 1962), is no longer able to grasp the complete picture of population migration in China because migrants with different socioeconomic statuses are facing totally

http://www.gov.cn/zhengce/content/2014-07/30/content_8944.htm). More details about this system will be presented in Chapter 3.

different situations and thus have differential propensities to migrate. Therefore, instead of the scale of migration, it is imperative to obtain more understanding about the skill composition of migrants and its underlying determining mechanism. In other words, understanding the question “Who has a higher propensity to migrate” is much more valuable and crucial than the question “how many people move” for current economic development and social stability. As Storper (2018) stated, “Rather than looking for interregional variation in average wage and housing price curves and possible points of overlap, these should be disaggregated for different skill groups.”

1.1.2. Who has a higher propensity to migrate?

To answer the question “who has a higher propensity to migrate,” it is necessary to realize the self-selection mechanism of migrants, which determines the skill composition of migrants. Because of different socio-economic statuses, people obtain heterogeneous returns from migration, which in turn influences their migration motivations and decisions, and finally reshapes their aggregated migration probabilities at the macro level. In other words, migration is inherently a selective process on different characteristics, especially the skill level. Generally speaking, a positive selection of migrants occurs when high-skilled individuals have a higher propensity to migrate than low-skilled individuals and vice versa for a negative selection of migrants. This research particularly focuses on the educational self-selection of migrants since their education levels significantly impact the economic growth of origins and destinations (as discussed previously). As such, the first central question of this thesis is:

- *What is the self-selection pattern of migrants in China?*

To answer this question, we need to compare the migration incentives of individuals with different education levels. Based on four periods of Census data from the *National Bureau of Statistics* (NBS), Figure 1.3 depicts the migration scale (hist) and rate (line) of interprovincial migrants by educational levels, demonstrating the migration patterns of different education groups.

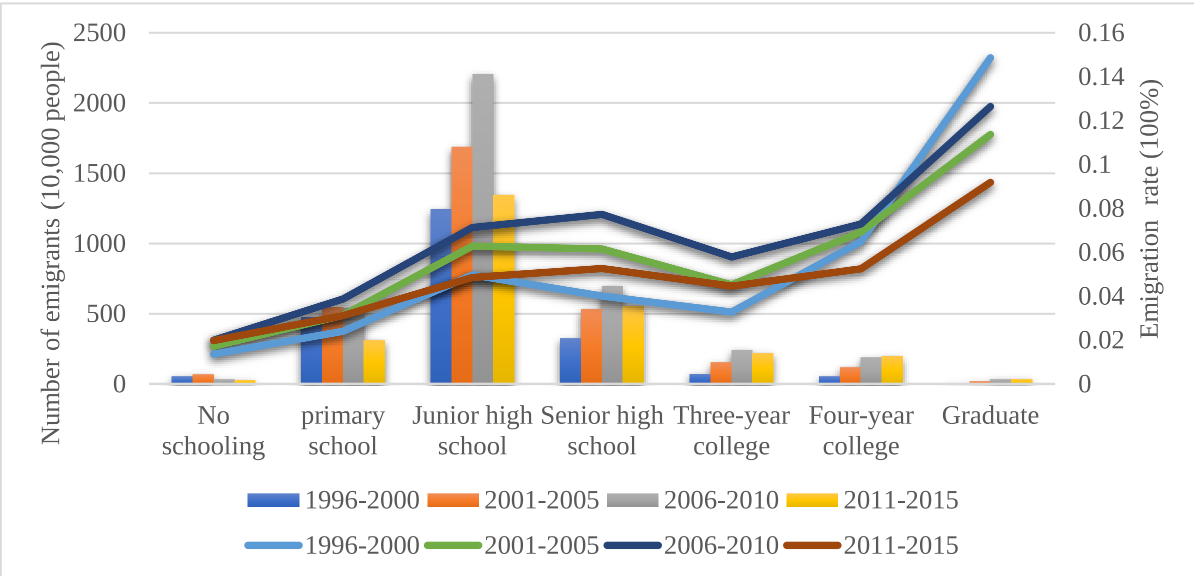

Figure 1.3 Interprovincial migration quantities and rates across education levels, 1996-2015

Note: the data comes from Census data, 2000, 2005, 2010, 2015[2].

[2] NBS only provides subsamples of 2000, 2005, 2010, 2015 census microdata source, which accounts for 0.95‰, 2‰, 0.95‰ and 1‰ of total population, respectively. These data sources do not contain income or other skill information, so the skill level is simply represented by the education level. Only interprovincial migration with five-year interval is considered in this analysis, as the four databases all defined migrants as those whose usual residence is different from 5 years ago at a provincial scale. Emigration quantities is estimated by the migrant number of each Census and their representation degrees, while emigration rates are calculated by the ratio of number of migrants to that of all population.

With the largest share of the total population, people with a high school degree also make up the largest share of migrants, while the quantities of low- and high-educated migrants are far less. This explains the left-skewed inverted U-shape of education distribution of migrant amounts in the histogram chart of Figure 1.3. In contrast, the emigration rate across education levels reveals a different pattern, as shown in the Line chart of Figure 1.3. The emigration rate is increasing non-monotonically with the education level, while individuals with three-year college degrees surprisingly have relatively lower migration rates than their counterparts. Actually, after controlling disturbances of demographics and regional fixed effects, the inter-provincial migration in China has a U-shaped selection pattern of migrants (as shown in Figure 3.4), implying that mid-educated people have the lowest propensities to migrate than other counterparts (please see detailed discussions in Section 3.3).

This U-shaped selection pattern is completely different from cases of international migration or internal migration in other countries. International migration scholars have identified a variety of migrant selection patterns, including positive selection (Grogger & Hanson, 2011), negative selection (Abramitzky et al., 2012; Borjas, 2008), or intermediate selection (Chiquiar & Hanson, 2005; Gould & Moav, 2016). In contrast, positive selection is more common in internal migration, as found by many urban economics scholars (Behrens et al., 2014; Combes et al., 2012; De la Roca, 2017). However, none of these cases reveals a similar selection pattern to that of inter-regional migrants in China. Why does China have such a particular pattern of migrant selection? This question necessitates an in-depth investigation of the impact mechanism behind this pattern.

1.1.3. Why does China have such a particular pattern of migrant selection?

Regional income differentials and migration costs are commonly the two most important factors to explain population migration in the narrative of neoclassical economics (Harris & Todaro, 1970; Sjaastad, 1962; Todaro, 1969). Yet, while providing a good explanation of macro-level migration scale or individual-level migration decisions, they do not shed light on the skill composition of the whole migrant group. For example, why are some regions losing more highly skilled talents while others are losing more low-skilled labor? Similarly, why do some regions attract more high-skilled talents while others have an influx of more low-skilled migrant workers? These questions reveal the self-selection pattern of migrants that traditional migration analysis cannot explain. To answer these questions, we need first to clarify the mechanism influencing the differences in migration returns between migrant groups before we can understand the differences in their actual migration behavior. This leads to the second central question of this thesis:

- *What is the underlying mechanism inducing the particular self-selection pattern of migrants in China?*

This thesis proposes four potential dimensions to answer this question:

Firstly, the *hukou* system, as a unique internal migration policy, still exists and affects all aspects of migrants' lives. As shown in Figure 1.2, there are still many migrant workers who have not obtained a local *hukou*, meaning that they are not effectively protected by the local welfare system. However, the *hukou* system is always skill-biased because local governments employ it as a policy tool to attract targeting talents and exclude the "low-end (*Di Duan*)" population. As a result, this system has formed a vast selection system for migrants, which profoundly influence migration returns for migrants with different skill levels and thus their propensities to migrate. Nevertheless, despite intense debate about other aspects of this system, the impact mechanism of the *hukou* system on migrant selection is hardly discussed in the literature.

Secondly, along with the regional development disparities, there are also significant differences in the returns to skills across regions. It has been documented comprehensively that coastal regions experience a greater increase in skill premiums (measured by return to education) due to globalization (Han et al., 2012), industry specialization (Li, 2018), and the increasing relative supply of skilled labor (Zou et al., 2009). This skill premium should lead more high-skilled individuals to migrate from inland to coastal regions to maximize their income levels. However, this mechanism still remains unknown in China. Besides, return to education only explain a portion of income, and there are many more components that go unexplained. It is also interesting to investigate whether other income components (in fact, their variations) affect migrant selection patterns.

Thirdly, in addition to the skill premiums changes, the labor market structure is also changing since the skill-biased technological change. Responding to the general policy of industrial upgrading in China, digitization and automation are rising significantly, but there are substantial regional differences. The resulting impact on the regional labor market structure has been dramatic since the technological change has always been friendly only to highly skilled groups but destroys the economic gains of low-skilled groups (Acemoglu & Restrepo, 2020; Autor et al., 2003). Of course, this impact will finally spill over to migrant workers because of their considerable share of the labor market. Whether will this process influence the self-selection of migrants? Unfortunately, we know little about this question.

Fourthly, the rising social issues because of economic development may exert heterogeneous effects on different migrant groups. Unlike the environmental issues where all will suffer almost the same adverse effects, the unaffordable housing costs seem to place an additional burden on the low-skilled group, thus affecting their economic returns from migration. In the meantime, the local public housing system turns away migrants without a local *hukou*, further aggravating their housing conditions (Shi et al., 2016). Will this unaffordability crowd out low-skilled and low-income migrants? Previous studies do not give a clear answer, and more in-depth studies are needed.

1.2. Research questions

Given that the self-selection of migrants is vital to regional development but received limited attention in China, this research aims to investigate the mechanism underlying the educational self-selection of migrants in China, which involves analyzing migrant selection patterns and its underlying mechanism when the specific condition of China is considered. Accordingly, this research aims to answer two central questions:

- What is the self-selection pattern of migrants in China?
- What is the underlying mechanism inducing the particular self-selection pattern of migrants in China?

The first question focuses on the selection pattern of migrants across regions, while the second core question concerns the underlying mechanism of the pattern. This research proposes four dimensions discussing the self-selection of migrants, which raises four sub-questions:

- How does the *hukou* system reshape the selection pattern of migrants in China?
- How do different income components influence the migrant selection pattern in China?
- How does the technological change influence the skill composition of migrants in China?
- How do unaffordable housing costs select migrants in China?

1.3. Research objectives

According to the proposed research questions, this research aims to achieve five research objectives:

1. To demonstrate the selection pattern of migrants in China.

2. To uncover the underlying mechanism of migrant selection in China.

    2.1 To investigate the role of the *hukou* system in reshaping migrant selection.

    2.2 To analyze the relationship between different income components and migrant selection.

    2.3 To explore the labor market consequences of technological changes on migrants' skill composition.

    2.4 To reveal the heterogeneous crowding-out effects of unaffordable housing costs on migrants with varying skill levels.

1.4. Chapter layout

This research consists of five chapters to achieve the five research objectives:

Chapter 2 first gives a brief review of related literature and proposes the research framework of this thesis.

Regarding the first objective, which is to analyze the selection pattern of migrants in China, this research first employs the Census data to demonstrate the migration rate[3] by education

[3] Migration flows cannot capture the migration incentives because of the interference of its population base. For example, the population with secondary education is the largest proportion of the total population, which leads to

levels, which captures migration incentives across education levels. Then, the binary logit model is further used to eliminate the interference of personal demographic characteristics and illustrate the difference in migration probabilities. These works will be conducted in chapter 3.

After identifying the selection pattern of migrants, this research further discusses the underlying mechanism. Firstly, in chapter 3, this research incorporates the *hukou* system into the self-selection framework to theoretically analyze its impact mechanism on migrant selection. Then, a conditional logit model is used to discuss the *hukou* system's two proposed impact channels empirically. These works answer how the *hukou* system influence migrant selection and achieves objective 2.1.

Secondly, this research discusses the heterogeneous effects of two different income inequality, inequality of opportunity and effort, on migrant selection in chapter 4. The income inequality will be first decomposed into two parts based on a Mincer-type estimation procedure. Then, the two components will be introduced into the macro-level empirical self-selection model to discuss their effects on migrant selection patterns captured by the migration rate difference between low- and high-educated individuals. The empirical results can answer the different roles of two income inequality components, achieving objective 2.2.

---

the largest number of migrants. As such, the migration rate, calculated by the ratio of migrants to the population with same education level, is more suitable to capture migration incentives given that higher migration incentives promote higher proportions of population migration across education levels.

Thirdly, this research discusses how technological change influences migrant selection in chapter 5. Since there are many technological change dimensions, this chapter takes one specific case, industrial robot installation in different regions, to demonstrate this effect. A Bartik-type robot exposure index is developed to capture the robot exposure condition in different cities. Then, the micro-level empirical self-selection model is employed to estimate the robot exposure index's effects on the skill levels of migrants. By doing so, the question of how technological change affects migrant selection can be answered from the small entry point of industrial robot installation, which fulfils objective 2.3.

Finally, this research turns to the role of housing costs in migrant selection in chapter 6. The housing costs are first introduced into the basic self-selection framework to conduct theoretical analysis and propose several research hypotheses. After this, the same micro-level empirical self-selection model as in chapter 5 is employed to verify these proposed hypotheses. These works can answer how unaffordable housing costs select migrants and achieve objective 2.4.

In summary, all objectives have been achieved in the above four chapters. Chapter 7 then concludes all significant results. This chapter also discusses practical implications for regional and national governments relying on these concluding remarks.

1.5. Significance of research

This research is significant in both theory and practice. Beyond empirical studies, this research extends the self-selection framework of migration with four impact channels that have not been discussed before. Based on the extended framework, this research theoretically and empirically investigates the self-selection of migrants in China as well as the underlying mechanism, which answers the two central research questions in China.

Regarding the theoretical perspective, this research has several contributions to the literature. First, beyond "push and pull" theories, this research provides a new angle to observe the population migration in China. A complete picture of how migrants with varying skill levels behave under heterogeneous socio-economic conditions will be given in this research, which is particularly valuable in today's knowledge economy, where human capital is increasingly imperative. Second, this research discovers some new channels through which the self-selection of migrants is impacted, including the unique *hukou* system in China, inequality of opportunity, technological change, and housing costs, which are hardly discussed in previous studies. Third, based on these new channels, this research has extended the self-selection framework into the scope of internal migration. This new framework bridges the knowledge gap of differences between internal migration and international migration in terms of the underlying mechanism, such as differences in return to skills, migration costs, and selective migration policies.

Regarding the practical perspective, this research also provides some significant implications for government planners at both national and regional levels. A better understanding of differences in migration behavior among heterogeneous populations provides local governments with insights into effective tools for attracting more high-skilled migrants and thus boosting regional economic development. In the meantime, the uneven distribution of human capital may also result in regional development disparity, which is harmful to economic development at the national level and needs more attention from government planners. Furthermore, this study also highlights the unfair conditions faced by migrants with different socioeconomic statuses in the shadow of the selective *hukou* system, social inequality, technological change, and soaring housing prices. The national and regional governments can obtain perceptions from these results, reconsider the *hukou* reform and rising social issues, and adopt differentiated policy tools to improve the living conditions of varying skill groups.

CHAPTER 2. RESEARCH FRAMEWORK

2.1. Migration theories and self-selection of migrants

Under the narrative of neoclassical economics, population migration can be explained by the utility (income) maximization framework. One will migrate only if his/her net utility/income gain is larger than zero after netting migration costs. From a macro perspective, the geographic differences in the supply and demand of labor influence the regional income differentials, which motivates people to migrate from low-income to high-income regions (Harris & Todaro, 1970; Lewis, 1954). This process will gradually adjust the labor market of two regions and eventually reach an equilibrium. In addition to income, migration costs induced by regional differences, such as urban amenities and housing prices, will also influence the process of equilibrium (Roback, 1982; Rosen, 1979). From a micro perspective, the cost-benefit calculation based on the utility maximization framework for each individual determines their personal migration choices (Sjaastad, 1962). One can choose the destination from multiple choices to maximize their migration returns given his/her characteristics. In the meantime, they also need to undertake migration costs, such as transportation costs, assimilation costs of different cultures or languages, and new skill learning costs.

This framework under neoclassical economics is the most prevalent tool to explain migration among previous studies. However, they fail to explain the underlying mechanism of the self-selection of migrants, given that the macro-level framework focuses only on the average effects of macro-level factors to describe the migration size, while the micro-level framework pays attention only to the personal decisions of individuals. Explaining the self-selection of migrants involves not only differences in patterns among migrant groups, but also the unequal economic and social returns to the different socioeconomic characteristics behind them. As such, a new theoretical framework is needed, which not only captures the heterogeneity among individuals but also the composition of all migrants at the macro level.

Based on the Roy (1951) model and neoclassical economics framework, Borjas (1987) proposed a self-selection framework to explain the skill composition of migrants. Starting from the most straightforward two-region migration, the same group of residents living in region 0 earn wage ω_0 and ω_1 in region 0 and 1, respectively (the subscript indicating an individual is ignored). Accordingly, the wage gain I from region 0 to region 1 netting migration cost C is:

$$I = \log \omega_1 - \log(\omega_0 + C) \approx \log \omega_1 - \log \omega_0 - \pi, \tag{2.1}$$

where π is a "time-equivalent" measure ($\pi = C/w_0$). According to Roy (1951), the wage level in two regions can be decomposed into two parts: the mean wage μ and wage variation ηs dependent on the regional return to skills η and personal skill level s[4], such that, $\log \omega_0 = \mu_0 + \eta_0 \mathrm{s}$ and $\log \omega_1 = \mu_1 + \eta_1 \mathrm{s}$. Similarly, the migration cost is also

[4] In relation to internal migration, the labor market is more unified, meaning that skill level is identical across all regions.

heterogeneous across skills and can be decomposed into mean costs μ_π and cost variation $\eta_\pi s$, such that $\pi = \mu_\pi + \eta_\pi s$. Therefore, the wage gain I can be extended as:

$$I \approx \log\omega_1 - \log\omega_0 - \pi = (\mu_1 - \mu_0 - \mu_\pi) + (\eta_1 \mathrm{s} - \eta_0 \mathrm{s} - \eta_\pi s). \tag{2.2}$$

Equation 2.2 reveals that wage gain I is determined by two components, one unrelated to the personal skill level and the other related. Since wage gain I determines the migration incentives for each individual, the derivative of wage gain I with respect to skill level s gives an intuitive indication of the difference in migration incentives across different skill levels, such that:

$$\frac{\partial I}{\partial s} = \eta_1 - \eta_0 - \eta_\pi. \tag{2.3}$$

From Equation 2.3, we can infer the selection pattern of migrants. If $\frac{\partial I}{\partial s} > 0$, there is a positive selection of migrants, meaning that high-skilled individuals have relatively higher wage gains from migration and thus have more migration incentives than low-skilled individuals, and vice versa if $\frac{\partial I}{\partial s} < 0$.

In addition, this equation indicates that two key factors determine the self-selection pattern of migrants: difference in return to skills $(\eta_1 - \eta_0)$, and heterogeneous migration costs (η_π). On the one hand, more unequal regions with higher returns to skills $(\eta_1 > \eta_0)$ attract more high-skilled migrants (i.e., positive selection), while low-skilled migrants prefer more equal regions with lower returns to skills $(\eta_1 < \eta_0)$ (i.e., negative selection). On the other hand, migrants undertake heterogeneous migration costs influencing their migration returns. Previous studies usually assume that migration costs decrease with increasing skill levels ($\eta_\pi < 0$) because high-skilled migrants have sufficient information, better assimilation ability, and fewer policy restrictions (Belot & Hatton, 2012; Chiquiar & Hanson, 2005; McKenzie & Rapoport, 2010). The resulting heterogeneous migration costs have asymmetrically prevented low-skilled migrants from migrating, leading to positive or intermediate selection patterns.

This theoretical framework was tested by multiple international migration cases, such as Mexico-US migration (Chiquiar & Hanson, 2005), Puerto Rico-US migration (Borjas, 2018), Israel-US migration (Gould & Moav, 2016), and other cross-countries migration in OECD countries (Beine et al., 2011; Grogger & Hanson, 2011). However, the self-selection of migrants in internal migration has hardly been discussed, and limited studies focus mainly on the selection pattern rather than the inducing mechanism behind it.

Generally speaking, the positive selection of migrants is the most common pattern of internal migration in most cases. In the literature branch of urban economics, scholars have identified the positive selection of migrants to megacities and take it as the driving force of agglomeration economics (Behrens et al., 2014; Combes et al., 2012; Roca & Puga, 2017). However, they do not delve into the factors behind this positive selection, except to argue that skill returns are higher in megacities (Borjas et al., 1992; De la Roca, 2017). In contrast, migration literature explains this positive selection by the inherently lower migration costs of high-skilled migrants but does not further explore the asymmetry of such migration costs (Bauernschuster et al., 2014). Therefore, more research is still needed to discuss why high-skilled individuals are more mobile.

Regarding the internal migration in China, Du et al. (2005) employed two household survey data in China's poor regions to study the migration of poor households and identified an inverted-U shaped relationship between household endowments and migration possibilities (i.e., intermediate selection). In Contrast, Wu (2010) finds a positive selection pattern that young, male, and better-educated individuals with good health in rural areas have higher migration possibilities, i.e., a positive selection. Xing's (2014) mixed results show that permanent rural-urban migrants are positively selected, while the selection pattern for temporary rural-urban migrants is unclear. On the one hand, mixed results of these studies implies the complicated selection pattern in China, which may be inconsistent with cases found in other countries. On the other hand, rural-urban migration is the main focus of these studies, with little attention paid to migrant selection regarding inter-regional migration, which may show a complete different selection pattern.

In summary, three research gaps are warranting more investigations and discussions. First, current studies regarding the self-selection of migrants mainly focus on international migration and neglect internal migration. Given higher accessibility and fewer restrictions, the features of internal migration may be distinct from those of international migration, questioning the usefulness of the classical self-selection framework. Second, the triggering mechanisms behind the typical positive selection of internal migrants have hardly been adequately studied. There should be deeper reasons leading to higher skill returns and lower migration costs for high-skilled individuals and thus the higher mobility of them. Third, the self-selection pattern of internal migrants in China may be inconsistent with cases in other countries, but it is still understudied, such as the self-selection of inter-regional migrants and the factors behind the induced selection. This research aims to contribute to these three research gaps by investigating the self-selection of inter-regional migrants in China. The following section will further discuss the underlying mechanism of migrant selection in China.

2.2. The mechanism underlying the self-selection of migrants in China

As documented above, returns to skills and heterogeneous migration costs are two key factors inducing migrant selection. However, there are deeper impact mechanisms that shape the influence of these two superficial factors, such as different regime systems and migration diasporas in international migration. This research proposes three new underlying factors inducing migrant selection regarding internal migration: inequality of opportunity, technological change, and housing costs. In addition, selective migration policies also play an essential role in migrant selection, which will be discussed in this research by taking the unique *hukou* system as a case study. Notably, this section only provides a brief discussion, while detailed theoretical discussions are developed in each corresponding chapter.

2.2.1. Return to skills

2.2.1.1. Inequality of opportunity

Return to skills is the key driving force inducing migrant selection, which has received sufficient discussions in international migration. Yet, how to define the return to skills is a core question influencing the final selection pattern. Previous studies have considered numerous measures capturing skills, including overall wage (Belot & Hatton, 2012; Grogger & Hanson, 2011), predicted wages denoting observable skills (Chiquiar & Hanson, 2005), residual wages denoting unobservable skills (Borjas et al., 2019), and education (Gould & Moav, 2016). In such cases, the skill selection pattern of migrants varies with returns to different dimensions of skills. For example, Gould and Moav (2016) find that the emigration rate from Israel increases with education due to higher return to education in the US. However, the selection pattern regarding residual wages is inverse U-shaped because of different returns to unobservable skills captured in residual wages.

These studies share a common idea that a certain type of skill selection pattern can only be influenced by the economic return to this specific type of skill because the latter directly determines the wage gain from this type of skill. Accordingly, the difference in returns to education between origin and destination determines the education selection pattern of migrants. Differing from this idea, this research proposes that education selection may also be influenced by other components of inequality through indirect channels.

Generally speaking, personal income levels can be influenced by personal efforts (e.g., personal choices on education, occupation, and working/learning hours) or their endowed social opportunities (e.g., gender, race, and family background). Political philosophy scholars recognized the outcome inequality induced by uneven social opportunities as "inequality of opportunity (IOP)" and the rest caused by varying degrees of personal efforts as "inequality of effort (IOE)" (Rawls, 1971; Sen, 1980). These two components have been identified to have significant but opposite effects on human capital formation and economic performance (Chiu, 1998; Marrero & Rodríguez, 2013; Mejía & St-Pierre, 2008; Song et al., 2020; Song & Zhou, 2019). Considering that migration will also influence human capital accumulation, it is interesting and crucial to investigate whether inequality of opportunity will influence population migration and, thus, human capital accumulation.

This research proposes that two inequality components influence the education selection of migrants through direct and indirect impact channels, respectively. On the one hand, IOE is highly correlated with return to education and thus influence the education selection of migrants directly. In other words, we can assume that lower levels of inequality of effort in destination increase migration incentives more for low-educated migrants. On the other hand, IOP damages the economic returns of individuals with poor circumstances (or less social opportunities), such as female, minority, and poor family background (Barber, 2000; Durand & Massey, 1992; Martin, 2007). Noting that these disadvantages will simultaneously undermine their access to education resources and result in lower education levels (Gamboa & Waltenberg, 2012; Palomino et al., 2019), IOP actually crowds out individuals with low education levels. Furthermore, the low-educated may also chase an equal society with more social opportunities to benefit their offspring through migration. In summary, this indirect impact mechanism leads to a similar assumption to IOE that lower levels of inequality of opportunity in destination increase migration incentives more for low-educated migrants.

2.2.1.2. Labor market shock induced by technological change

Moreover, the regional return to skills difference may no longer reflect the differences in the national wealth distribution system but the supply and demand for different skill types. The demand for high or low-skilled labor, induced by labor market differences across cities, constitutes the underlying reason for the different migration patterns of high and low-skilled migrants (Diamond, 2016). As a result, the labor market condition should largely influence the regional returns to skills and thus the self-selection pattern of migrants in internal migration.

The most significant factor that has reshaped the labor market in the past few decades is technological change. There have been heated discussions and debates about the labor market effects of technological change since the last century (Acemoglu & Restrepo, 2020; Autor et al., 1998; Murphy et al., 1998). However, most of these studies focus on the whole labor market and resulting economic performance, while few of them paid attention to its impact on population migration, let alone the selection and sorting of migrants. This raises the research question remaining unknown that how the technological change alters the local labor market and thus reshapes the selection pattern of migrants.

Generally speaking, technological change exerts two effects on the labor market. On the one hand, technological change will promote economic growth and industry expansion, create more jobs, and thus increase the demand for labor, i.e., the *reinstatement effect* (Acemoglu & Restrepo, 2018, 2019). On the other hand, the technological change is usually skill-biased, implying that jobs created usually require highly skilled employees to fill them. In contrast, low-skilled (in fact, mid-skilled) routine-type jobs are gradually substituted by automation and flow to the service sector requiring lower skill levels, i.e., the *displacement effect* (Autor et al., 2003; Autor & Dorn, 2013; Goos & Manning, 2007). These two forces act together in the labor market, leading to a polarization of skills and wages in the labor market.

In the meantime, these effects will also spread to migrant workers, which is hardly discussed in the literature. This question is particularly important in China, where migrants contribute significantly to the labor force (Freeman, 2015). Given migrants' inherently low skill levels and less social security in destination cities, the skill-biased technological change may have hit them (especially low-skilled ones) more severely than locals. To confirm these concerns, this research takes the installation of industrial robots as an entry point to discuss the consequences of the technological change to migrants. According to the above analysis, regions with more industrial robot installation (i.e., higher degrees of robot exposure) tend to attract more high-skilled migrants and crowd out low-skilled ones from the local production sector. These displaced migrants may flow to local service sectors or regions with lower robot exposure. As such, technological change will reshape the selection pattern of migrants significantly.

2.2.2. Heterogeneous migration costs

In addition to income incentives, the migration cost is also an essential factor affecting selection when correlated with the skill level (Borjas, 1991). Previous studies regarding international migration have attempted to explain heterogeneous migration costs by various factors, including poverty constraints (Belot & Hatton, 2012), migration networks (Beine et al., 2011; McKenzie & Rapoport, 2010), immigration policies (Bertoli et al., 2016; Haddad, 2020), and pre-migration skill training (Jaschke & Keita, 2021).

Differing from migration costs in international migration, internal migration costs are usually too small to limit migration significantly. Concerning China, the highly developed railway network ensures low physical transportation costs, while the high degree of cultural unity lowers the psychological cost of migration. However, the high degree of urbanization in coastal regions has also triggered various problems. Economic development has increased people's income levels but at the same time has also driven up prices, especially housing prices.

Previous studies have confirmed that housing plays a significant role in migrants' settlement decisions (Yang & Guo, 2018), welling being (Li & Liu, 2018), and local social integration (Wang & Fan, 2012). Regarding the migration decision, some empirical results seem to come to the counter-intuitive conclusion that soaring housing prices did not deter migrants from moving (Chen et al., 2019; Wu et al., 2019). Yet, this argument will be challenged when looking at different migrant groups. Zhang et al. (2016) argued that the gradually widening income gap caused the increase in housing prices, which in turn increased the housing cost burden for middle- and low-income groups. Chen et al. (2019) found the attraction of megacities for high-skilled individuals is gradually decreasing with time due to housing prices. Unfortunately, they do not provide direct evidence of the heterogeneous effects of housing costs on different migration groups. The current research aims to fill this research gap and investigate how housing unaffordability induces heterogeneous migration costs and thus select migrants.

Through a theoretical analysis (please see Section 6.3 for detailed theoretical modelling), this research proposes that even though housing prices adversely affect all migrants, this effect is relatively greater for low-skilled groups. As a result, the barriers built by housing costs mainly crowd out low-skilled migrants. In other words, high housing costs at the destination discourage the migration of low-skilled migrants more significantly than high-skilled migrants, i.e., a position selection. Furthermore, different social groups may face different housing costs, such as different gender, age cohorts, and employment sectors. For example, state-owned enterprises will provide higher provident funds for house purchasing or directly provide housing after meeting certain conditions, which are rare in private enterprises. This heterogeneity results in various selection degrees by housing costs among different migrant groups.

2.2.3. Selective migration policies

Migration policy always plays a significant role in migration, especially international migration. As a convenient but effective way to control the quality of immigrants, an increasing number of governments introduced screening policies to select essential immigrants based on observable skills (*e.g.*, education levels and skill certificates) (Bertoli et al., 2016; de Haas et al., 2016). Numerous studies have theoretically studied the role of selective migration policies in migrant selection, suggesting that legal barriers have induced more migration costs for low-skilled migrants and thus resulted in intermediate or positive selection (Beine et al., 2011; Clark et al., 2007; Mayda, 2010).

Nevertheless, these studies lack empirical evidence on the one hand and pay no attention to internal migration on the other hand. The latter may be due to the fact that policy constraints within a country nearly do not exist and thereby exert no significant effects on migration in developed counties. This may not be the case in China, as there is a unique policy restriction, namely the household registration (*hukou*) system. Differing from international migration restrictions, the *hukou* system may reshape the migrant selection pattern by affecting both regional returns to skills and migration costs.

The consequence of the *hukou* system is not a new topic in academia but still very hot and attention-grabbing. Some scholars have proposed that migration in China is still not exhausted because of this system, hindering urbanization and agglomeration economies (Au & Henderson, 2006; Chan & Zhang, 1999). In addition, numerous studies have studied the other dimensions of the *hukou* system' consequences on migration, including migration scale (Bao et al., 2011), migrants' living conditions (Huang et al., 2010; Hui et al., 2014; Song & Smith, 2021; Tao et al., 2014, 2015), and migrants' labor market behaviors (Song, 2016; Zhang, 2010). However, how the *hukou* system influences the self-selection of migrants remains under-investigated.

Origin from 1957, the *hukou* system binds people's civil rights to their place of birth utilizing a *hukou* identity, while rigorous restrictions are imposed on the transfer of *hukou* location by the government. As a result, unlike the international selective immigration policies, the *hukou* system does not restrict people's free movement since the 1980s; instead, it profoundly influences the economic returns and costs of living in destination cities. On the one hand, this system limits migrants' access to local amenities, such as compulsory educational resources, medical resources, and social security (Song, 2014). On the other hand, migrants without local *hukou* will face significant labor market discrimination that undermines their earnings (Gravemeyer et al., 2011; Zhang, 2020). Moreover, the *hukou* system itself is a vast selective mechanism that favors highly skilled or wealthy labor. In a nutshell, the *hukou* system has formed a hierarchic system selecting migrants across regions through distorting return to skills by wage discrimination and causing heterogeneous migration costs by *hukou*-related amenities. As a result, the selection of migration in China stands somewhere between international migration and internal migration within developed countries.

2.2.4. Research framework

In summary, as key determinants of migrant selection, return to skills and heterogeneous migration costs continue to play essential roles in the self-selection of internal migrants, but through different impact channels from international migration. Regarding return to skills, in addition to return to education, education selection may also be influenced by other components of inequality through indirect channels. This thesis proposes two different income components, inequality of opportunity and inequality of effort, both of which affect educational self-selection. Then, beyond different components, this thesis further discusses the factors influencing return to skills, that is, labor market shock induced by technological changes, given that the returns to skills no longer reflect the equality degree of national distribution system but the supply and demand of labor in the regional market regarding internal migration.

Regarding heterogeneous migration costs, in addition to those found in international migration, internal migration has its own unique migration costs, that is, housing costs. This is particularly important in China since the sky-rocketing housing prices and worsening housing affordability in many cities. Therefore, this thesis will analyze the heterogeneous effects of housing costs on different migrant groups. Last but not least, unlike typical selective migration policies in international migration that affect migrant selection only through migration costs, China has its specific selective migration policy, the *hukou* system, which affects both the return to skills and the cost of migration. As a result, the *hukou* system significantly reshapes the self-selection pattern of migrants in China.

Figure 2.1 demonstrates the research framework of this thesis that uncovers the mechanism underlying the self-selection of migrants in China. To verify this research framework, this thesis employs four chapters discussing four new impact factors. The *hukou* system will be first discussed theoretically and empirically in Chapter 3 due to its significant role in migration. Then, the following three chapters will discuss inequality of opportunity, technological change, and housing costs, respectively.

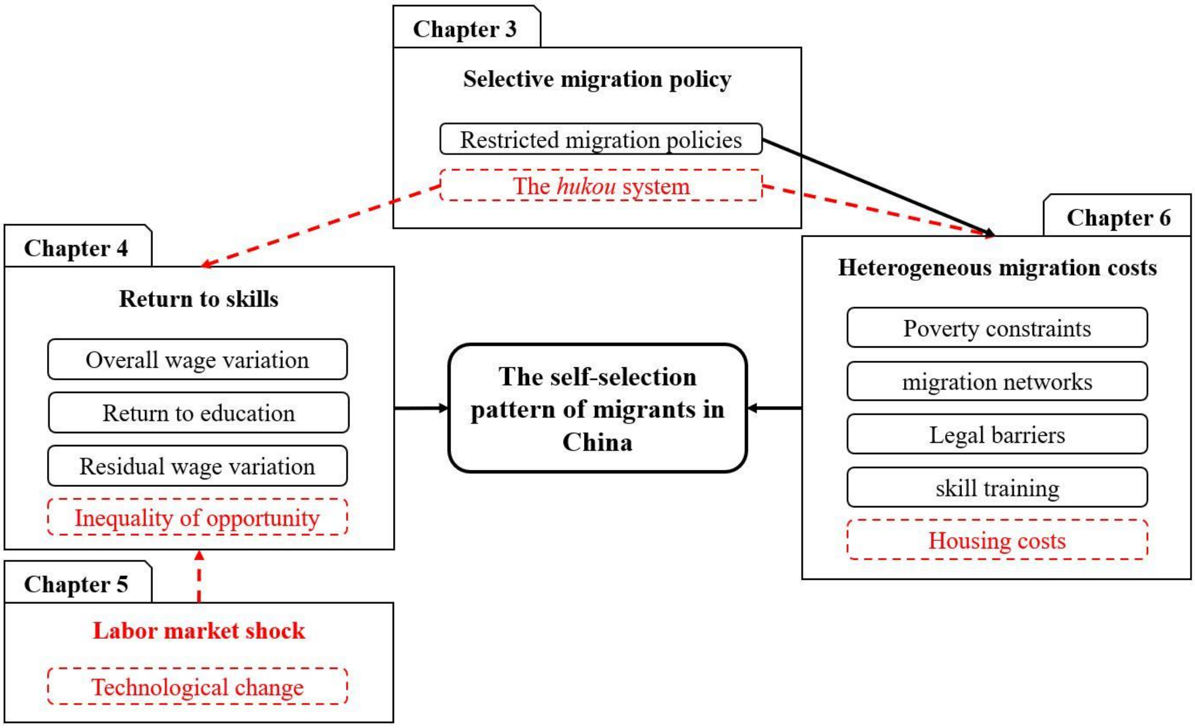

Figure 2.1 Research framework

Note: research gaps are marked in red font or dashed lines.

CHAPTER 3. THE SELECTION PATTERN OF MIGRANTS IN CHINA AND THE ROLE OF THE *HUKOU* SYSTEM

3.1. Introduction

This chapter aims to discuss the educational self-selection pattern of migrants in China and how the *hukou* system influences this pattern theoretically and empirically, and thus achieve Objectives 1 and 2.1. To achieve these goals, this chapter first carefully checks the self-selection pattern of migrants using a binary logit model to investigate differences in migration probabilities between education levels. Then, this chapter incorporates the *hukou* system into the self-selection framework to conduct theoretical modelling. This analysis proposes two impact channels of the *hukou* system: distorting return to skills and limiting hukou-related amenities for non-local migrants. Finally, this chapter uses the conditional logit model to examine factors influencing migrants' destination choices, which identifies the proposed two impact channels of the *hukou* system.

The rest of this chapter is structured below: Section 3.2 gives a comprehensive literature review of self-selection of migrants and the *hukou* system in China; Section 3.3 demonstrates the migrant selection pattern; Section 3.4 proposes a theoretical self-selection framework incorporating the *hukou* system and models the selection pattern of migrants in China; the proposed impact channels are discussed in Section 3.5; the final Section 3.6 concludes this chapter.

3.2. Literature review

3.2.1. Selective migration policy and self-selection of migrants

In international migration, migration policies play an important role in the migrant selection by skills. Since 1945, international migration policies have evolved from restriction to selection, given that most restriction policies only prohibit irregular migrants and family migrants instead of high-skilled migrants or students (de Haas et al., 2016). The primary purpose of this turning is to compete for more high-skilled labor to accelerate the knowledge economy. For this aim, countries have proposed various types of selective migration policies to target these talents, such as the Canadian "human capital" model, the Australian "neo-corporatist" model, and the market-oriented, demand-driven model (Koslowski, 2014). Consequently, it could be expected that selective policies have intensified the positive selection of immigrants, but there are limited studies on this topic.

Theoretically, several studies have introduced selective policies into the Roy model to analyze their effects on migration patterns (Beine et al., 2011; Clark et al., 2007; Mayda, 2010). The general thinking is that legal barriers have induced more migration costs for low-skilled migrants and thus resulted in intermediate or positive selection. By contrast, only a few studies have assessed the consequences of these selective policies empirically. From the spatial dimension, Czaika and Parsons (2017) compared the differential effects of several skill-selective policies on the scale of high-skilled migrants and the skill composition of all migrants in 10 OECD destinations. Their results demonstrate that the points-based systems increase the number of high-skilled migrants and intensify the positive selection in the meantime. From the temporal dimension, Haddad (2020) identified that the evolution of migration policies reshaped migrant selection patterns from negative to positive after 1982 regarding migration from French Overseas departments to metropolitan France. These pieces of evidence have revealed the significant direct effects of selective migration policies on migrant selection patterns through restrictions to low-skilled migrants and incentives to high-skilled counterparts.

In comparison, selective migration policies are less common in internal migration. The positive selection is a typical pattern for migrants to big cities (Behrens et al., 2014; Combes et al., 2008). For instance, De la Roca (2017) and Bacolod et al. (2021) both found a significantly positive selection of migrants to big cities in Spain and Colombia, respectively. These studies discuss migrant selection on the premise of free movement without legal barriers. In such cases, the migrant selection pattern is dominated by market factors instead of institution factors, such as return to skills and human capital agglomeration.

Nevertheless, this premise may not apply in China in the presence of the *hukou* system. Although it did not discuss migrant selection, Bao et al. (2011) empirically identified the adverse effects of *hukou* restrictions on the scale of migration. Xing (2014) investigates the self-selection pattern of rural-urban migration in China. His results demonstrate different selection patterns between permanent migrants (with local *hukou*) and temporary migrants (without local *hukou*), which partially suggests *hukou*'s vital role in migrant selection. In other words, the analysis of inter-regional migrant selection in China cannot bypass the discussion of the *hukou* system.

This chapter aims to provide unique evidence on how selective migration policies influence migrant selection regarding internal migration. The *hukou* system in China is taken as a particular case study for the discussion. Unlike migration policies in international migration, the *hukou* system in China influences migrant selection through indirect channels instead of direct restrictions. However, to the best of our knowledge, no study has discussed the close relationship between the *hukou* system and the self-selection of inter-regional migrants in China. The following sub-section will comprehensively document how the *hukou* system has been deeply embedded in China's population migration over several decades.

3.2.2. The *hukou* system in China

The household registration (*hukou*) system, promulgated in 1958, divided China's population into rural and urban *hukou* bounded by their birthplace and parents, which largely influences the internal migration in China. Before the reform and opening up in 1978, a person from rural areas and wanted to be a permanent urban resident needed two steps: i) convert *hukou* status from agricultural/rural to non-agricultural/urban (*nongzhuanfe*i), and then ii) change the location of *hukou* registration. In this period, the central government dominated this process and allocated limited *nongzhuanfei* quotes to exceptional applicants each year, preventing nearly all population migration.

In the early 1980s, China has liberalized the free movement of the population but does not include *hukou* transfer. Instead, the central government devolved the *hukou* system management to local governments. This reform enables many cities to abolish *nongzhuanfei* and promulgate local *hukou* policies to settle more permanent or semi-permanent migrants, especially in the late 1990s. However, to some extent, these reforms may have made the permanent settlement for migrants in destination cities harder than before because local governments have raised thresholds for local *hukou* obtainment and formed a selective system to attract targeted migrants (Chan & Buckingham, 2008). On the other hand, these reforms all concentrate on the abolition of *nongzhuanfei*, while the *hukou* location transfer barrier still exists and impedes inter-regional migration. Consequently, inter-regional migrants benefit little from these reforms while the transfer of *hukou* location is still challenging and skill-based.

Generally speaking, there are four channels to obtain a local *hukou* in other regions: investment, housing purchase, talent program, and employment (Zhang et al., 2019). Migrants have to either invest a high quantity of wealth (*e.g.*, housing) in destination cities or obtain a qualified high level of skill (through education or skill certificates) to obtain a local *hukou*. If these conditions are not met, migrants need to spend more time (usually several years) and money (for social security fees). As shown in Figure 3.1, the proportion of migrants who obtained a local urban *hukou* among all interprovincial migrants significantly increases with their education levels. Moreover, the threshold is increasingly higher with the size of cities (Zhang et al., 2019). As a result, the *hukou* system has formed a sorting mechanism based on migrants' wealth and skill levels, resulting in spatial hierarchies within China. This mechanism produces a unique group named "floating population" (*i.e.*, migrants without a local *hukou*), who account for the largest proportion of migrants (over 90% of interprovincial migration flows).

Along with the localized management of the *hukou* system, the financial and welfare systems were also devolved to local governments. As a result, although benefiting from the large number of labor inflow, destination regions do not need to provide corresponding social amenities to all migrants because of the separation of movement freedom and citizenship (Chan, 2009). In other words, the *hukou* relaxation was not equivalent to the free population movement in other countries. Under the *hukou* system, one can freely migrate to other regions but cannot enjoy some local services and welfare benefits unless he/she obtains a local *hukou* through any of the channels mentioned above.

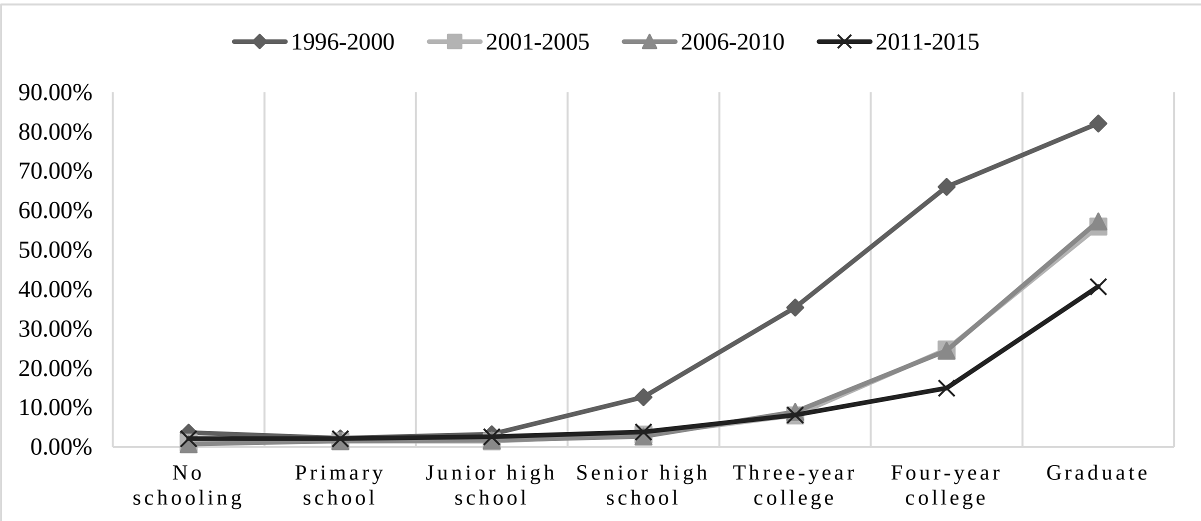

Figure 3.1 The proportion of migrants who have obtained a local urban *hukou* among all interprovincial migrants

Notes: 1. the proportion is calculated by authors based on four Census data; 2. please see Section 3.4.2 for the definition of migrants.

Even under these restrictions, China still experienced rapid economic development and witnessed the "Great Migration," particularly the migration from inland regions to developed coastal regions (Cao et al., 2018; Fan, 2005; Liang, 2001; Liu & Shen, 2014b; Shen & Liu, 2016). Related data show that 35% of China's total workforce in 2015 was accounted for by migrants, a higher level than the share of migrants in the total Chinese population (Freeman, 2015).

There have been many documents discussing how the *hukou* system affects migrants, which is mainly reflected in two aspects: social amenities and economic returns. Firstly, because of the binding of social amenities to the *hukou* location, migrants are excluded from the welfare system of the destination city. In relation to education resources, which are traditionally crucial for Chinese families, migrant children cannot access the same level of compulsory education as locals in some big cities (Qian & Walker, 2015; Zhang, 2017). As a result, only adult people migrate for economic opportunities, leaving the elderly and children in the countryside without care (Chang et al., 2011; Ye & Lu, 2011). Another essential dimension is that social security programs, including social relief, social welfare, and social insurance, are also not applicable for non-local migrants (Song, 2014). Faced with the absence of these social amenities, migrants finally have to return home or keep moving to other regions, especially for low-skilled interprovincial migrant workers (Wang & Fan, 2006; Zhang et al., 2020).

Secondly, the vast income gap between regions cannot be thoroughly enjoyed by migrants due to *hukou*-induced wage discrimination. A pile of literature has proved the existence of wage discriminations between local workers and non-local workers without *hukou* (Gravemeyer et al., 2011; Song, 2016; Zhang, 2020). Fu and Ren (2010) find that the difference in return to education between natives and migrants is the key channel that the *hukou* system influences migrants' labor market economic returns. Moreover, the degree of discrimination varies with *hukou* types, sectors, and income levels. Song (2016) performed a simple analysis based on quantile regressions, which showed that the 50% and 75% quantiles of migrants in the wage distribution suffer most from wage discrimination, while low-income migrants are hardly discriminated against and even more dominant in the labor market. In other words, low-income migrants are more likely to find a relatively equivalent or better job than locals with an observationally equivalent skill level, while the opposite is true for mid- or high-income migrants.

In summary, unlike the international selective immigration policies, the *hukou* system does not directly restrict people's free movement; rather, it profoundly influences the economic returns and social amenities of migrants in destination cities. In the meantime, the *hukou* system is a huge selective mechanism that favors highly skilled or wealthy labor. Consequently, the migration in China is likely to stand somewhere between free internal migration in developed countries and restricted international migration, which supports the applicability of international migration theory (i.e., self-selection framework) in China's internal migration.

3.3. The selection pattern of migrants in China

3.3.1. Empirical model and data

This section attempts to analyze the selection pattern of inter-regional migrants. A binary logit model is employed to measure the difference in migration probabilities across skill levels, as shown below:

$$logit(Mig_p) = \boldsymbol{edu\_dum_p \alpha} + \boldsymbol{X_p \gamma} + \varepsilon, \tag{3.1}$$

where Mig_p indicates the migration decision of individual *p*, equals to one when choosing to migrate and otherwise zero; $\boldsymbol{edu\_dum_p}$ is a vector of dummy variables indicating different education levels of individual *p*, which can capture the non-linear migration probabilities across education levels; $\boldsymbol{X_p}$ is a vector of personal characteristics, including age, gender, marital status, and occupations. In this equation, the coefficient vector $\boldsymbol{\alpha}$ directly captures the migration probability differences between different education levels and thus demonstrates the general selection pattern.

This research obtains population migration data from the *China National Bureau of Statistics*'s (NBS) population census and survey data for four periods: Population Census data in 2000 and 2010 and 1% population sampling survey data in 2005 and 2015. This research takes individual-level sub-samples from these four databases, which represent 0.95‰, 2‰, 0.95‰, and 1‰ of all populations in four periods, respectively. According to the survey setting, this research defines migrants as individuals whose current residence is different from their residence five years prior. This method can only analyze population movements at the provincial level, but it contains all migrants (with or without local *hukou* at destination) in the last five years. This chapter mainly considers economic migration. In other words, the observations should be in the employment markets. Consequently, this research excludes individuals under 20 and over 60 or who did not have a job. After selection, the final sample has 589,376, 1,241,039, 629,293, and 613,422 observations in 2000, 2005, 2010, and 2015, respectively. In addition to the migration related data, census data of four periods also provide sufficient information to help eliminate the effects demographic characteristics influencing migration decision, including education levels, age, gender, Marital status, occupation type, and industry sector. All these variables will be controlled in the empirical model and their description is listed in Table 3.1.

Table 3.1 Description of independent variables

Variable	Descriptions	Data source
Personal characteristics		
Schooling years	Individuals' years of schooling (year)	Census data
Age	Individuals' age (year old)	Census data
Gender	Dummy variable, equal to one when the individual is male	Census data
Marital status	Dummy variable, equal to one when the individual is unmarried	Census data
Income		
Urban average income	The average disposable income per capita of urban people (Yuan)	China Yearbook 2011
Return to education	The economic return to years of schooling, manually calculated.	CFPS 2011
Migration costs		
Distance	The Euclidean distance between provincial capitals (km)	Manual calculation
Same province	Dummy variable, equal to one if the destination choice is the same as the origin.	Census data
Amenities		
Urbanization rate	The ratio of urban population to all population (100%)	China Yearbook 2011
Industry structure	The GDP ratio of Tertiary industry to Second industry (100%)	China Yearbook 2011

Student/teacher ratio	The ratio of student number to qualified teacher number (100%)	China Yearbook 2011
Doctor/resident ratio	Number of qualified doctors per 10000 people (person)	China Yearbook 2011
Medical insurance rate	The proportion of urban people participating in urban basic medical insurance at the end of the year (100%)	China Yearbook 2011
Urban greening	The area of urban park and green area per capita (m2/person)	China Yearbook 2011
Urban public transportation	Number of Public transport vehicles per 10,000 people (vehicles)	China Yearbook 2011

3.3.2. Migrant selection patterns

The discussion firstly focuses on the general educational selection pattern of migrants in China. Based on the Census data for four periods, this research calculated the emigration rate across education levels and migration scenarios (the ratio of out-migrant numbers to population numbers), as shown in Figure 3.2 and Figure 3.3. Figure 3.2 shows that the emigration rate increases with education levels regarding all migration and inland-coastal migration but remains stable regarding coastal-inland migration. However, the trend of inland-coastal migration is not strictly monotonically increasing, given that migrants with college degrees have equivalent emigration rates to (even slightly lower than) migrants with high school degrees. Figure 3.3 further confirms the stability of this relationship over four periods.

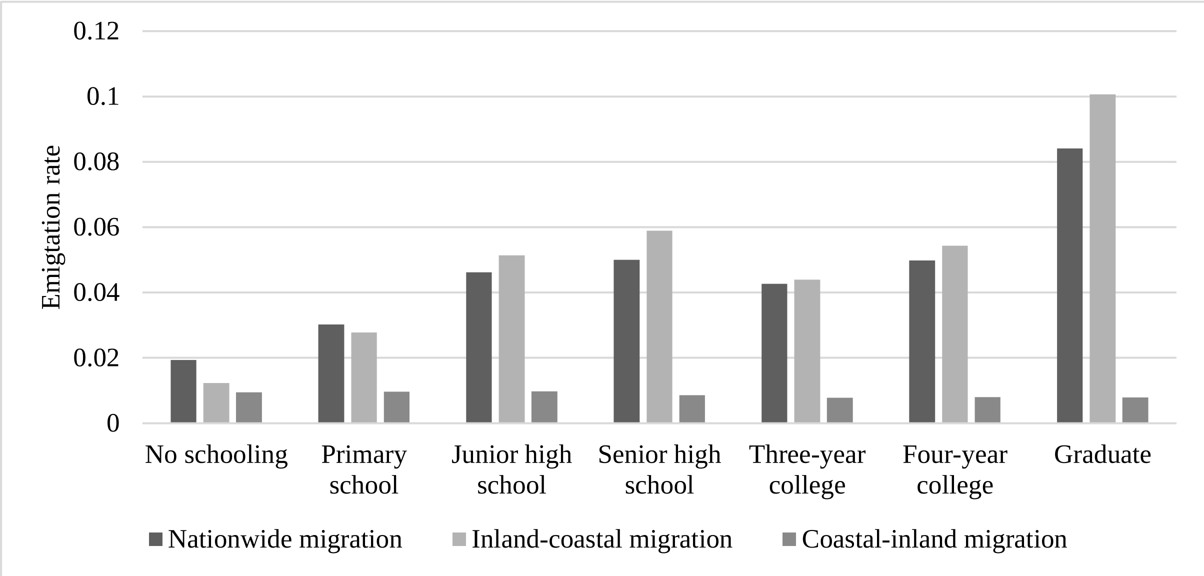

Figure 3.2 Emigration rate of different migration patterns during 2011~2015

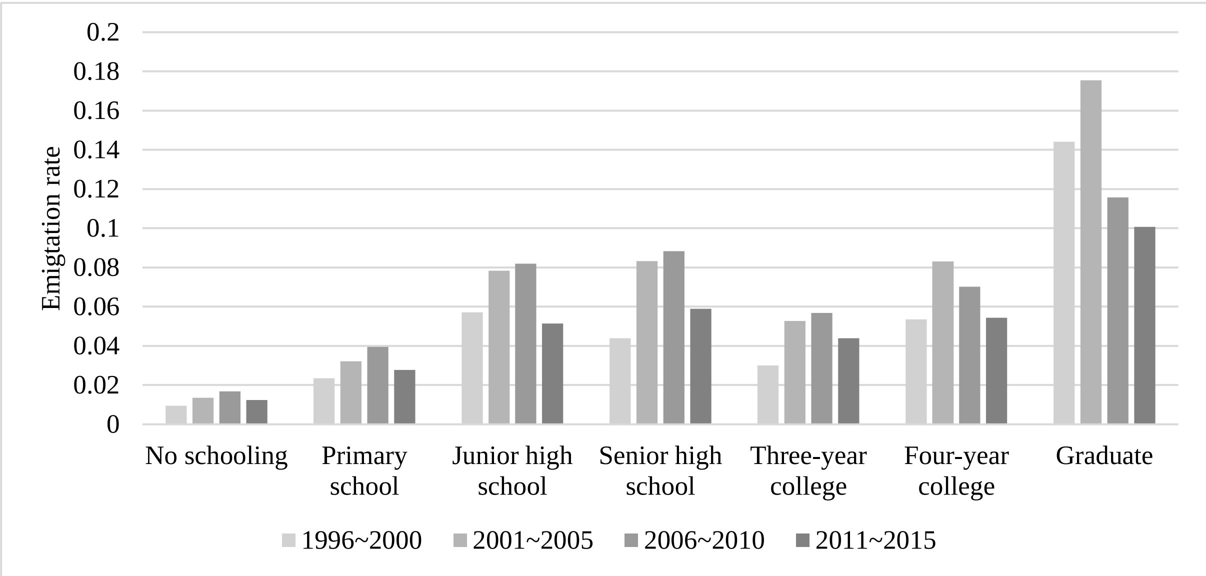

Figure 3.3 Emigration rate of inland-coastal migration during four periods

Table 3.2 Binary logit regression results of migration choice

Dependent variable: (Moving = 1)	Inland-Coastal					Nationwide	Coastal-Inland	Central-Coastal	Western-Coastal
	2011-2015	2011-2015	2006-2010	2001-2005	1996-2000	2011-2015	2011-2015	2011-2015	2011-2015
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Junior high school	0.314***	-0.410***	-0.418***	-0.495***	-0.400***	-0.424***	-0.397***	-0.675***	-0.062
	(0.026)	(0.029)	(0.023)	(0.018)	(0.027)	(0.022)	(0.066)	(0.038)	(0.043)
Senior high school	0.279***	-0.689***	-0.828***	-0.917***	-1.200***	-0.673***	-0.697***	-0.975***	-0.285***
	(0.029)	(0.033)	(0.028)	(0.022)	(0.036)	(0.026)	(0.080)	(0.043)	(0.052)
Three-year college	-0.103***	-0.951***	-1.344***	-1.251***	-1.335***	-0.823***	-0.816***	-1.083***	-0.924***
	(0.038)	(0.043)	(0.039)	(0.033)	(0.065)	(0.032)	(0.102)	(0.054)	(0.077)
Four-year college	0.124***	-0.592***	-0.954***	-0.547***	-0.292***	-0.471***	-0.713***	-0.716***	-0.578***
	(0.041)	(0.047)	(0.046)	(0.040)	(0.083)	(0.035)	(0.109)	(0.059)	(0.086)
Graduate	0.943***	0.349***	-0.072	0.841***	1.559***	0.448***	-0.663***	0.079	0.654***
	(0.099)	(0.107)	(0.106)	(0.100)	(0.212)	(0.066)	(0.233)	(0.127)	(0.201)
Age	-0.063***	-0.614***	-0.064***	-0.085***	-0.090***	-0.058***	-0.054***	-0.065***	-0.056***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.003)	(0.001)	(0.002)
Gender	0.213***	-0.088***	-0.183***	-0.257***	-0.289***	0.073***	0.281***	0.008	-0.282***
	(0.017)	(0.018)	(0.016)	(0.013)	(0.021)	(0.014)	(0.044)	(0.023)	(0.021)
Marital status	-0.334***	-0.381***	-0.419***	-0.695***	-0.657***	-0.258***	0.435***	-0.334***	-0.450***
	(0.021)	(0.023)	(0.205)	(0.018)	(0.027)	(0.018)	(0.065)	(0.029)	(0.039)
Cons	-0.821***	-3.163***	-3.789***	-3.360***	-3.139***	-1.200***	-2.697***	-2.812***	-3.334***
	(0.041)	(0.253)	(0.164)	(0.109)	(0.229)	(0.146)	(0.333)	(0.312)	(0.431)
Industry FE controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Occupation FE controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Origin FE controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	356,467	356,467	373,456	747,637	363,389	613,422	249,127	190,892	165575

Pseudo R^2	0.066	0.280	0.354	0.481	0.429	0.212	0.046	0.243	0.340

Notes: 1. Industry types include dummy variables indicating secondary and tertiary industries; Occupation types include eight common occupation categories according to Chinese occupation classification standard *GBT 6565-2015*. 2. Central provinces include Shanxi, Jilin, Heilongjiang, Anhui, Jiangxi, Henan, Hubei, Hunan; Western provinces include Inner Mongolia, Guangxi, Chongqing, Sichuan, Guizhou, Yunnan, Tibet, Shaanxi, Gansu, Qinghai, Ningxia, Xinjiang. 3. Standard errors are in parentheses; ***, **, and * denote significance at the 1, 5, and 10 percent levels, respectively.

The above preliminary results have roughly revealed the selection pattern of migrants but are disturbed by personal characteristics, such as different age cohorts, gender differences, and others. Therefore, this research further employs a binary logit model to estimate the emigration rate across education levels after controlling personal characteristics and fixed effects. The results are shown in Table 3.2.

Firstly, let us consider the first migration scenario (migrating from inland to coastal regions) without controlling fixed effects, as shown in Column (1) of Table 3.2. Except for the extremely high migration probability of individuals with *graduate* degrees, other groups have similar migration probabilities. Notably, the relationship between migration probability and education levels shows a U-shaped pattern, given that individuals with three-year college degrees have the lowest migration probabilities. This pattern is more significant after controlling industry, occupation, and origin fixed effects, as shown in Columns (2) of Table 3.2. For visualization, this research further plots coefficients of different education level dummies in Figure 3.4, which shows a clear-cut U-shaped selection pattern.

To ensure the stability of results, different periods are considered, and the U-shaped selection pattern still exists across three other periods, as shown in Columns (3) to (5) of Table 3.2. Then, this research divides inland regions into central and western regions and finds similar results, as shown in Columns (8) and (9) of Table 3.2. Accordingly, these results identify that a) there is a generally positive selection of all migrants and b) however, the selection pattern is not monotonically increasing but a U-shape.

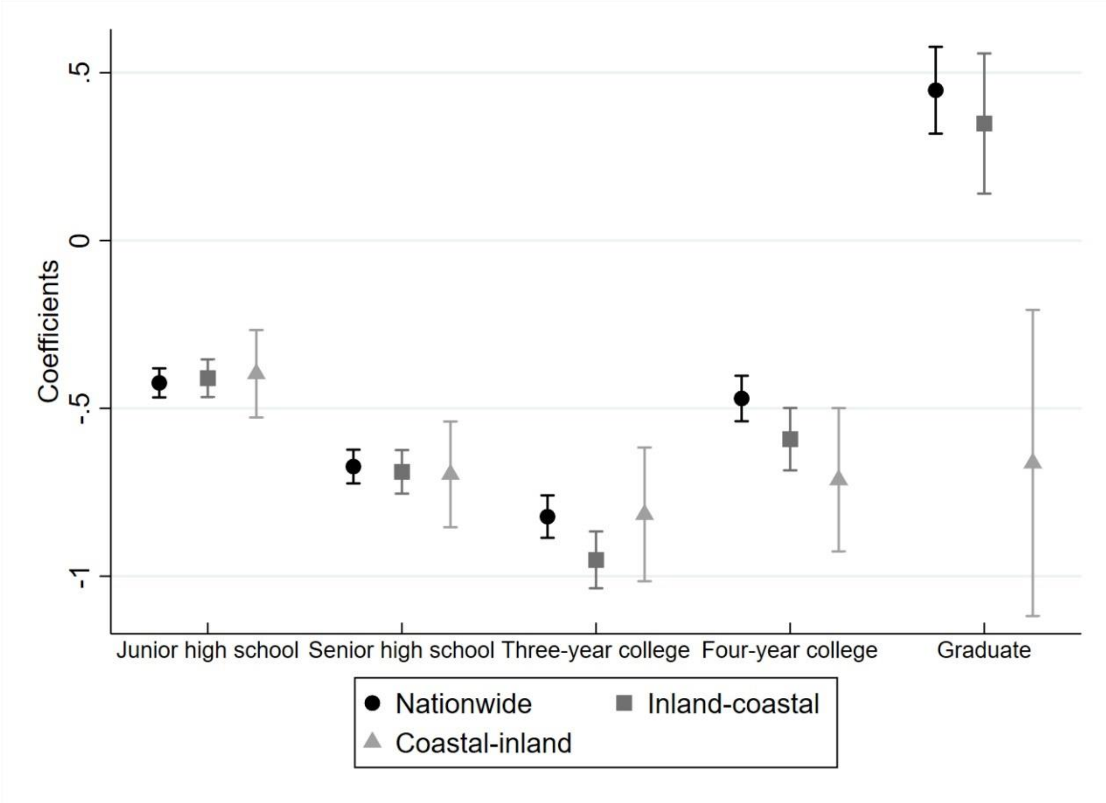

Figure 3.4 Migration probability of migrants during 2011-2015

Note: the probability takes migrants with primary school degrees or below as reference.

Secondly, this research further discusses the migration of other scenarios, as shown in Columns (6) and (7) of Table 3.2. Column (6) of Table 3.2 shows that the migrant selection pattern across all regions is similar to that from inland to coastal regions. The possible reason is that inland-coastal migration dominates the migration in China. Column (7) of Table 3.2 estimates the second migration scenario that individuals migrating from inland to coastal regions and demonstrates a negative selection of migrants from coastal to inland regions. As shown in Figure 3.4, the intensity of selection decreases with education levels.

In addition to education selection, the migration probability varies with other demographic characteristics. According to Table 3.2, young individuals have significantly higher migration probabilities than their old counterparts, given the significant negative coefficient of *age*. In the meantime, unmarried individuals have lower migration probabilities across all migration scenarios. Regarding gender, the coefficient varies with different migration scenarios. For coastal-inland migration and migration across all regions, male individuals have higher migration probabilities than female counterparts, but inland-coastal migration is just the opposite.

In summary, this subsection finds that these is clear U-shaped selection pattern in China, which is totally different from other migration pattern in both international and internal migration. Why does China have this pattern? What is the specific role of the *hukou* system? The following subsection propose a theoretical analysis to discuss the role of the *hukou* system in migrant selection in China, based on the self-selection framework.

3.4. Theoretical framework

3.4.1. Self-selection framework

To better identify factors influencing the selection pattern in China, this research conducts a simple theoretical analysis based on the self-selection framework proposed by Borjas (1987, 1991). Assuming there are multiple locations, individual p migrates from origin I to destination j and obtain different utility levels, as shown in the following utility functions:

$$U_{p,i} = \ln w_{p,i} + A_{p,i} \tag{3.2a}$$

$$U_{p,ij} = \ln(w_{p,j} - C_{ij}) + A_{p,j} \approx \ln(w_{p,j}) - \pi_{ij} + A_{p,j} \tag{3.2b}$$

where $U_{p,i}$ and $U_{p,ij}$ represent the utility functions of individual p in origin *I* and destination j, respectively; $w_{p,i}$ and $w_{p,j}$ are the wage levels of individual p at the origin *I* and destination j, respectively; C_{ij} measures the migration cost of individual p from origin *I* to destination j; π_{ij} is a "time-equivalent" measure ($\pi_{ij} = C_{ij}/w_{p,i}$) of migration costs; and $A_{p,i}$ and $A_{p,j}$ represent the amenities/disamenities at the origin *I* and destination j, respectively. The direct comparison between $U_{p,i}$ and $U_{p,ij}$ measures the gain from migration, and individual p will migrate to maximize $U_{p,ij} - U_{p,i}$.

The income gap is an essential driver of population migration, whether in China or abroad (Grogger & Hanson, 2011; Zhu, 2002). Because individuals have heterogeneous skill levels, they will obtain correspondingly different wage levels. As such, following Roy (1951), the multiple location wage equations can be written as follows:

$$\ln(w_{p,i}) = \mu_i + \eta_i s_p \tag{3.3a}$$

$$\ln(w_{p,j}) = \mu_j + \eta_j s_p \tag{3.3b}$$

where μ_i and μ_j are the mean wage of the same residents at the origin *I* and destination j, respectively; s_p is the individual p's skill level that is equivalent throughout the country; $\eta_i > 0$ and $\eta_j > 0$ refer to the return to skills at the origin *I* and destination j, respectively. Notably, these two equations measure wage levels of the same residents living at the origin *I* or migrating to the destination j to control skill differentials induced by demography structure difference.

In addition, this framework considers heterogeneous migration costs that vary with skill levels. In general, migration costs decrease with skill levels because high-skilled migrants tend to have better adaptability and information acquisition skills (Chiquiar & Hanson, 2005; McKenzie & Rapoport, 2010). As such, the migration costs can be written as follows:

$$\pi_{ij} = \mu_{\pi,ij} - \eta_\pi s_p \tag{3.4}$$

where π_{ij} is the migration cost from origin *I* to destination *j*; $\mu_{\pi,ij}$ is the mean migration costs of all migrants; $\eta_\pi > 0$ captures the general heterogeneous costs varying with personal skills.

3.4.2. The *hukou* system

This research introduces a key ingredient in this framework: the *hukou* system. Generally speaking, there are some skill thresholds for migrants to obtain local *hukou* in some big cities, such as the point system in *Beijing* and *Shanghai*. Accordingly, a skill threshold can be assumed for migrants that determines whether they can get the local *hukou* effortlessly. Migrants' *hukou* status can now be written as follows:

$$H_{p,j} = \begin{cases} 1 & if\ s_p > s_j^* \\ 0 & if\ s_p < s_j^* \end{cases} \tag{3.5}$$

where $H_{p,j}$ represents whether one can obtain a *hukou* effortlessly in region *j* (1 means yes); s_j^* captures the skill threshold of *hukou* obtainment in region *j*. The requirements to obtain a *hukou* vary across regions. The criteria in developed regions are continually rising over time to control the urban population, while other regions have set relatively lower constraints. The influence of the *hukou* system is reflected in two channels: social amenities and wage discrimination.

Firstly, the reason for the restrictions in *hukou* transfer lies in the linkage of *hukou* to some essential amenities in most cities. Therefore, amenities can be divided into non-*hukou*-related and *hukou*-related amenities, as follows:

$$A_{p,j} = a_j + H_{p,j} \cdot b_j \tag{3.6}$$

where a_j is the average non-*hukou*-related amenity level in destination j; and b_j captures the *hukou*-related amenities. $H_{p,j}$ was introduced to adjust the *hukou*-related amenity level of individual p. Highly skilled migrants ($H_{p,j} = 1$), who have a higher possibility to obtain a *hukou*, enjoy more *hukou*-related amenities, such as children's education, medical insurance, and others. The non-*hukou*-related amenity level was assumed to be region-specific and uncorrelated with skill level; each individual can enjoy these amenities such as urban greening and transportation.

Secondly, wage discrimination, which influences the return to skills of migrants, reflects the other channel of the *hukou* system's effects. According to Song (2016), mid- and high-skilled migrants (without local *hukou*) suffer from higher wage discrimination levels than low-skilled migrants. As such, it can be assumed that the wage discrimination level is negative correlated with skill levels, as follows:

$$J_{p,j} = \gamma_j(1 - H_{p,j})s_p \tag{3.7}$$

where $J_{p,j}$ measures the excepted wage discrimination level of individual p in destination j. Because the *hukou* status directly induces wage discrimination, $H_{p,j}$ was again introduced to adjust the wage discrimination level, which indicates that *hukou* discrimination only influence migrants who cannot obtain local *hukou* ($H_{p,j} = 0$). $\gamma_j > 0$ captures the level of wage discrimination that increases with skill levels. In such a case, the return to skills for migrants is no longer linear in the presence of wage discrimination.

3.4.3. Theoretical analysis of selection pattern

By summarizing the previous considerations, the complete utility functions finally come out:

$$V_{p,i} = \ln w_{p,i} + A_{p,i} = \mu_i + \eta_i s_p + a_i + b_i, \tag{3.8a}$$

$$V_{p,ij} = \ln(w_{p,j}) - \pi_{ij} + A_{p,j} - J_{p,j} = \mu_j + \eta_j s_p - (\mu_{\pi,ij} - \eta_\pi s_p) + a_j$$

$$+H_{p,j} \cdot b_j - \gamma_j(1 - H_{p,j})s_p. \tag{3.8b}$$

To maximize personal utility, one will migrate to region j only if

$$\Delta V_{ij} = V_{p,ij} - V_{p,i} = \mu_j - \mu_i - \mu_{\pi,ij} + a_j - a_i + H_{p,j} \cdot b_j - b_i$$

$$+(\eta_j - \eta_i + \eta_\pi + \gamma_j H_{p,j} - \gamma_j)s_p > 0. \tag{3.9}$$

According to migrants' skill levels, there are two conditions:

if $s_p > s_j^*$, $H_{p,j} = 1$, thus

$$\Delta V_{ij}{}^{h} = \mu_j - \mu_i - \mu_{\pi,ij} + a_j - a_i - b_i + b_j + (\eta_j - \eta_i + \eta_\pi)s_p \tag{3.10a}$$

$$\text{and } \frac{\partial \Delta V_{ij}{}^{h}}{\partial s_p} = \eta_j - \eta_i + \eta_\pi; \tag{3.10b}$$

if $s_p < s_j^*$, $H_{p,j} = 0$, thus

$$\Delta V_{ij}{}^{l} = \mu_j - \mu_i - \mu_{\pi,ij} + a_j - a_i - b_i + (\eta_j - \eta_i + \eta_\pi - \gamma_j)s_p \tag{3.11a}$$

$$\text{and } \frac{\partial \Delta V_{ij}{}^{l}}{\partial s_p} = \eta_j - \eta_i - \gamma_j + \eta_\pi, \tag{3.11b}$$

where $\Delta V_{ij}{}^{h/l}$ reflects the economic return to migration by skill levels; $\frac{\partial \Delta V_{ij}{}^{h/l}}{\partial s_p}$ captures the selection pattern of migrants. If $\frac{\partial \Delta V_{ij}{}^{h/l}}{\partial s_p} > 0$, there is a positive selection of migrants; otherwise, there is a negative selection.

Generally speaking, the internal migration in China shows an inland-coastal (west-east) pattern in the past decades because of the economic development gap between the coastal and inland regions, which is commonly documented in the literature (Cao et al., 2018; Fan, 2005; Liu & Shen, 2014b). Accordingly, this chapter mainly discusses two scenarios representing China's typical migration patterns: inland-coastal migration (scenario one) and coastal-inland migration (scenario two). In the meantime, scholars have also identified that coastal developed regions have a higher skill premium of wage (or return to skills) than inland regions (Asadullah & Xiao, 2019, 2020; Whalley & Xing, 2014; Zou et al., 2009). Therefore, it can be assumed that return to skills (η) in coastal regions is higher than that in inland regions.

Scenario one. Considering the migration from inland regions to coastal regions, return to skills in destination region is higher than that in the original region (i.e., $\eta_j > \eta_i$). Given that ${\Delta V_{ij}}^h - {\Delta V_{ij}}^l = b_j + \gamma_j s_p$, high-skilled migrants ($s_p > s_j^*$) have higher economic return to migration and thus get higher intention to move, which implies a positive selection for all migrants. For high-skilled migrants ($s_p > s_j^*$), there is a significant positive selection of migrants given that $\frac{\partial \Delta V_{ij}}{\partial s_p} > 0$. For low-skilled migrants ($s_p < s_j^*$), the selection pattern is ambiguous because the existence of wage discrimination (γ_j) lowers the return to skills of migrants. In a destination with a lower level of discrimination, $\eta_j - \eta_i - \gamma_j + \eta_\pi > 0$. In such a case, low-skilled migrants experience a low level of positive selection, as shown in (a) of Figure 3.5. In contrast, in a destination with a higher level of discrimination, $\eta_j - \eta_i - \gamma_j + \eta_\pi < 0$, which leads to a negative selection of low-skilled migrants, as shown in (b) of Figure 3.5.

According to the above analysis, there are mainly two inferences. Firstly, the higher return to skills in coastal regions will induce a significant positive selection of all migrants. In such a case, high-skilled migrants enjoy higher wages and better amenities, while low-skilled migrants obtain significantly worse amenities and suffer wage discrimination because of *hukou* barriers. Secondly, the *hukou* system distorts the monotonical selection pattern by undermining the return to skills of migrants through wage discrimination, which even leads to a U-shaped selection pattern. In other words, mid-skilled migrants ($s_p \rightarrow s_j^*$) might get the lowest returns from migration ((b) of Figure 3.5). On the one hand, they cannot enjoy the same level of amenities as high-skilled migrants; on the other hand, they suffer from severe wage discrimination than low-skilled migrants.

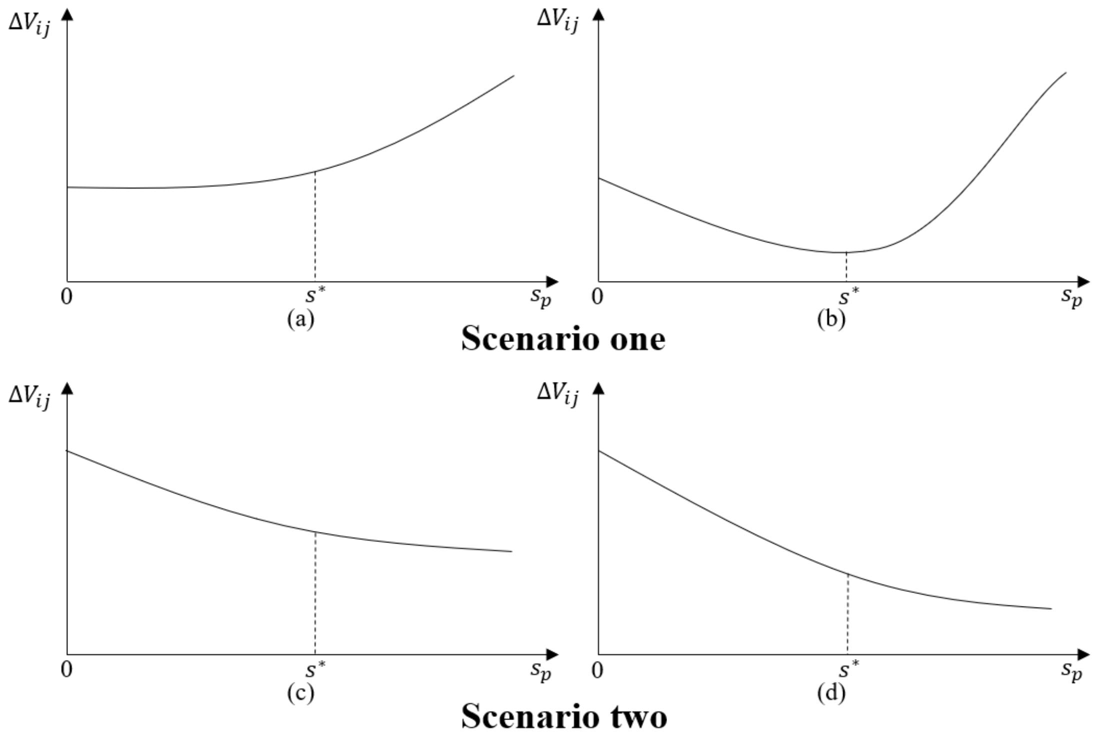

Figure 3.5 Selection patterns of two scenarios

Scenario two. Considering the migration from coastal regions to inland regions, return to skills in the destination should be lower, i.e., $\eta_i > \eta_j$. Given the lower return to skills, migrants from coastal to inland regions tend to be negatively selected. For high-skilled migrants ($s_p > s_j^*$), the sign of $\eta_j - \eta_i + \eta_\pi$ is uncertain given that $\eta_j - \eta_i < 0$, indicating that the negative selection of high-skilled migrants may not be significant. For low-skilled migrants ($s_p < s_j^*$), because of wage discrimination (γ_j), the negative selection should be more significant, and its degree increases with wage discrimination levels, as shown in I and (d) of Figure 3.5. Therefore, the inference from the above analysis is that the migration from coastal to inland regions generally has a negative selection, but the selection intensity decreases with skill levels.

In summary, these two scenarios, to a large extent, explain the migrant selection pattern in China and how the *hukou* system distorts this selection pattern. This theoretical model can also be applied to analyzing migrant selection patterns from small to big cities because the *hukou* obtainment criteria are also rising with city sizes (Zhang et al., 2019). Nevertheless, as discussed in this chapter, the Census data in China can only identify inter-provincial migrants in the recent five years. To verify the proposed two impact channels, the following subsection will further conduct an empirical to analyze the underlying mechanism of migrant selection in China.

3.5. Migrant selection channels

3.5.1. Empirical model and data

This subsection attempts to analyze the possible channels of the *hukou* system influencing the migrant selection pattern. Following Davies et al. (2001), this research employs the conditional logit model to study inter-regional migration. According to the theoretical framework, individuals will choose one region to maximize their utility (staying at original regions is also a choice). Assuming there are $\boldsymbol{J}$ regions for individuals to choose, the utility level that individual p can obtain in region j is:

$$V_{p,ij} = \boldsymbol{X_{p,ij}\beta} + \varepsilon_{p,ij}, \tag{3.12}$$

where $X_{p,ij}$ is a vector of region-specific attributes; $\varepsilon_{p,ij}$ captures random factors influencing the utility level. Following McFadden (1974) and assuming that $\varepsilon_{\boldsymbol{p},\boldsymbol{ij}}$ is independent and identically distributed with the Weibull distribution, the probability of migrating from region *I* to region *j* can be written as:

$$P_{p,ij} = P\left(V_{p,ij} > V_{p,ik} \middle| \forall k \in \boldsymbol{J}\right) = \frac{\exp \boldsymbol{X_{p,ij}\beta}}{\sum_{k\in \boldsymbol{J}} \exp \boldsymbol{X_{p,ik}\beta}}. \tag{3.13}$$

As discussed in Section 3.3, personal utilities after migrating are influenced by various factors, including wage levels, migration costs, and amenities. Accordingly, this research further specifies the $\boldsymbol{X_{p,ij}}$ as a set of variables:

$$\boldsymbol{X_{p,ij}} = \begin{Bmatrix} W_j, RTE_j, RTE_j \times edu_p, RTE_j \times {edu_p}^2, \\ \boldsymbol{D_{ij}}, \boldsymbol{D_{ij}} \times edu_p, \boldsymbol{A_j}, \boldsymbol{B_j}, \boldsymbol{B_j} \times edu_p \end{Bmatrix}, \tag{3.14}$$

where edu_p denotes individuals' schooling years; W_j denotes the average income level in region *j*; RTE_j is the return to education in region *j*; $\boldsymbol{D_{ij}}$ is a vector of variables capturing the migration costs between origin and destination regions; $\boldsymbol{A_j}$ and $\boldsymbol{B_j}$ are vectors of variables measuring non-*hukou*-related and *hukou*-related amenities, respectively. The interaction items between region-specific variables and schooling years can capture the heterogeneous effects of variables by individuals' education levels. In other words, interaction items reflect how these variables influence the migrant selection pattern in education.

This research can predict the coefficients of these variables according to theoretical inferences. Firstly, because of the assumption that migration costs decrease with skill levels, $\boldsymbol{D_{ij}} \times edu_p$ should have positive effects on destination choice. Secondly, as a consequence of wage discrimination, migrants cannot enjoy all returns to education in destination, the migrant selection by RTE_j may show a non-linear pattern (i.e., U-shaped pattern). As such, this research can predict that both $RTE_j \times edu_p$ and $RTE_j \times {edu_p}^2$ have significant impacts on destination choice, and their coefficient signs are negative and positive, respectively. Finally, the proposed theory implies that *hukou*-related amenities will also select migrants, predicting that $\boldsymbol{B_j} \times edu_p$ have significant impacts on destination choices, and their coefficient signs vary with amenity types.

The dependent variable is still using the migrant data derived from the Census data of four periods discussed in section 3.3.1. Regarding independent variables, three types are mainly considered: income, migration costs, and amenities. Firstly, in addition to *Urban average income*, it is needed to estimate the *Return to education*. Because census data do not cover the income information of individuals, this research utilizes the 2010 wave of *China Family Panel Studies* (CFPS) survey data to measure the return to education by provinces. This database is a national survey conducted by Peking University, covering 25 provincial administrative regions in China (excluding Hong Kong, Macao, Taiwan, Xinjiang, Tibet, Qinghai, Inner Mongolia, Ningxia, and Hainan). Based on these data, a standard Mincer wage equation is used to estimate the return to education by provinces (regressing incomes on years of schooling, age, square of age, gender, marital status, and *hukou* status).

Secondly, migration costs are highly correlated with migration distance. As such, the Euclidean *Distance* between provincial capitals can capture migration costs. Because the conditional logit model also considers the original region, the empirical model includes a dummy variable *same province* to indicate staying at the origin or not.

Thirdly, the empirical model further introduces several variables to capture amenities. Regional *Urbanization rate* and *Industry structure* characterize the employment market amenities. Because different urbanization rates or industry structures imply different demands for skill types, these variables will influence the migrant selection pattern. Then, the *hukou* system directly linked education, medical, and insurance welfares with the *hukou* location. Consequently, the empirical model introduces *Student/teacher ratio*, *Doctor/resident ratio*, and *Medical insurance rate* to capture these *hukou*-related amenities. Finally, the empirical model introduces the *Urban green area* and *Urban public transportation* to capture non-*hukou*-related amenities because they are public for everyone. The description of all independent variables is shown in Table 3.1.

3.5.2. Empirical results

This subsection discusses empirical results identifying possible impact channels of the *hukou* system on the migrant selection pattern. As shown in Table 3.3, conditional logit regression results give us a picture of how individuals choose the destination. This research firstly considers all populations, including natives and migrants. Because of the software's computing power limitation (nearly 1 million observations), only 7% observations of all samples in 2015 are taken. The four model specifications in Columns (1) to (4) of Table 3.3 show that pseudo R^2 all reaches 0.919, which means that variables have a high explanation power for destination choice. Then, this research only considers migrants' destination choice, as shown in Columns (5) to (6). By contrast, the Pseudo R^2 is reduced to about 0.324, implying a relatively lower explanation power.

Regarding independent variables, nearly all variables are statically significant at 1% level except for some interactions. As predicted, income levels captured by *Urban average income* and *Return to education * Schooling years* have significant positive effects on destination choice, as shown in Columns (1) of Table 3.3. In other words, individuals prefer destinations with higher income levels. The significant negative and positive coefficients of *Distance* and *Same province* reveal that individuals prefer staying at origin or destinations near origins because of migration costs. Moreover, regions with better amenities, captured by *Urbanization rate*, *Student/teacher ratio*, *Doctor/resident ratio*, *Medical insurance rate*, *Urban greening*, and *Urban public transportation*, are more attractive than others for all populations. By contrast, given the coefficient sign of *Industry structure*, more population still locate in regions with more Secondary industries.

The *hukou* system's impact is mainly reflected in selection on return to education and *hukou*-related amenities. Regarding the first channel, the interaction of square of schooling years and return to education is employed to capture the non-linear selection pattern induced by the regional return to education. For all populations, the significant coefficient of the interaction between schooling years and return to education in Column (1) of Table 3.3 reflects the positive selection of individuals' education level by the return to education. Then, Column (2) of Table

3.3 introduces the interaction of square of schooling years and return to education into the model but finds no significant effects on destination choice, revealing that there is no non-linear selection pattern of all individuals by the return to education. This research also considers different education level dummies and find a significant positive selection by the return to education, as shown in Column (3) of Table 3.3.

Nevertheless, for only migrants, the results show an entirely different pattern. Columns (5) and (6) of Table 3.3 show that return to education still positively selects migrants but not a monotonical selection pattern. The coefficients of *Return to education * Schooling years* and *Return to education * Square of schooling years* have opposite signs, implying a U-shaped selection pattern of migrants induced by the return to education. The model specification using education level dummies further verifies this result in Column (7) of Table 3.3, where migrants with senior high school degrees have lower migration probabilities than their counterparts.

Regarding the second channel, more interactions are further introduced into the empirical model in Columns (4) and (8) of Table 3.3. In terms of migration costs, the coefficient of *Distance * Schooling years* is significantly positive for all populations but not significant for migrants. This means that higher educated individuals are more likely to migrate, but the distance will no longer select migrants' education levels after deciding to migrate. In terms of amenities, the educational selection by amenities is not significant except for the *Student/teacher ratio* and *Doctor/resident ratio* regarding all populations, as shown in Column (4) of Table 3.3. By contrast, for migrants only, the educational selection by amenities is more significant. In detail, *hukou*-related amenities (including *Student/teacher ratio*, *Doctor/resident ratio*, and *Medical insurance rate*) positively select migrants, and all coefficients are significant at 1% level, as shown in Column (8) of Table 3.3.

The above results demonstrate the different conditions between natives and migrants. As discussed in Section 3.3, the *hukou* system should pay much responsibility for these results. Firstly, the *hukou* system distorts the educational selection by returns to education through wage discrimination. As a result, the educational selection by returns to education is not monotonically increasing but a U-shaped pattern for migrants instead of a positive selection for all populations. Secondly, *hukou*-related amenities enlarge the amenity level differences across education levels, leading to a more significantly positive educational selection for migrants. In summary, these results support main theoretical inferences in Section 3.3 and can statistically reflect how the *hukou* system influences the migrant selection pattern in China.

Table 3.3 Conditional logit regression results of migration destination choice during 2011~2015

Dependent variable: destination choice	Natives & migrants				Migrants			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Income								
Urban average income	1.911***	1.911***	1.918***	1.930***	3.165***	3.159***	3.156***	3.064***
	(0.255)	(0.255)	(0.255)	(0.255)	(0.144)	(0.144)	(0.144)	(0.144)
Return to education	-23.388***	-18.072***	-12.807***	-18.562***	-14.158***	6.065**	-8.078***	8.613**
	(3.136)	(6.336)	(2.677)	(5.171)	(1.378)	(2.790)	(1.235)	(3.445)
*Return to education * Schooling years*	1.781***	0.698		1.274***	0.625***	-3.395***		-3.550***
	(0.252)	(1.156)		(0.478)	(0.106)	(0.492)		(0.534)
*Return to education * Square of schooling years*		0.050				0.184***		0.184***
		(0.053)				(0.022)		(0.023)
*Return to education * Junior high school dummy*			5.166**				0.241	
			(2.462)				(1.060)	
*Return to education * Senior high school dummy*			11.020***				-4.112***	
			(2.780)				(1.197)	
*Return to education * Three-year college dummy*			14.078***				3.273**	
			(3.351)				(1.496)	
*Return to education * Four-year college dummy*			18.621***				11.470***	
			(3.573)				(1.555)	
*Return to education * Graduate dummy*			35.609***				18.237***	
			(7.586)				(3.126)	
Migration costs								
Distance	-0.002***	-0.002***	-0.002***	-0.002***	-0.001***	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Same province	5.674***	5.673***	5.672***	5.677***				

	(0.063)	(0.063)	(0.063)	(0.063)				
*Distance * Schooling years*				0.000**				0.000
				(0.000)				(0.000)
Amenities								
Urbanization rate	3.779***	3.786***	3.775***	5.548***	7.719***	7.707***	7.700***	13.426***
	(0.579)	(0.579)	(0.579)	(1.434)	(0.324)	(0.324)	(0.324)	(0.793)
Industry structure	-1.122***	-1.120***	-1.120***	-0.871***	0.191***	0.188***	0.187***	0.201*
	(0.076)	(0.076)	(0.076)	(0.231)	(0.038)	(0.038)	(0.038)	(0.110)
Student/teacher ratio	0.264***	0.264***	0.264***	0.157***	0.317***	0.317***	0.317***	0.274***
	(0.013)	(0.013)	(0.013)	(0.033)	(0.007)	(0.007)	(0.007)	(0.014)
Doctor/resident ratio	0.021***	0.021***	0.021***	0.001	-0.037***	-0.037***	-0.037***	-0.071***
	(0.003)	(0.003)	(0.003)	(0.010)	(0.001)	(0.001)	(0.001)	(0.005)
Medical insurance rate	1.625***	1.623***	1.627***	0.445	2.034***	2.046***	2.055***	-2.505***
	(0.336)	(0.336)	(0.336)	(0.978)	(0.239)	(0.240)	(0.240)	(0.589)
Urban greening	0.024***	0.023***	0.023**	0.025***	-0.101***	-0.101***	-0.101***	-0.100***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.006)	(0.006)	(0.006)	(0.006)
Urban public transportation	0.158***	0.158***	0.158***	0.157***	0.316***	0.317***	0.317***	0.319***
	(0.017)	(0.017)	(0.017)	(0.017)	(0.006)	(0.006)	(0.006)	(0.006)
*Urbanization rate * Schooling years*				-0.181				-0.544***
				(0.127)				(0.069)
*Industry structure * Schooling years*				-0.025				-0.007
				(0.021)				(0.010)
*Student/teacher ratio * Schooling years*				0.010***				0.005***
				(0.003)				(0.001)
*Doctor/resident ratio * Schooling years*				0.002**				0.003***
				(0.001)				(0.000)

*Medical insurance rate * Schooling years*				0.125 (0.090)				0.437*** (0.050)
Observations	39,233*25	39,233*25	39,233*25	39,233*25	24,791*25	24,791*25	24,791*25	24,791*25
Pseudo R^2	0.919	0.919	0.919	0.920	0.324	0.324	0.325	0.327

Notes: 1. Standard errors are in parentheses; 2. ***, **, and * denote significance at the 1, 5, and 10 percent levels, respectively.

3.6. Conclusions and discussions

This chapter studies the self-selection pattern of migrants in China and how selection migration policies influence it using China's unique household registration (*hukou*) system as an entry point, which provides an interesting and meaningful case for restricted internal or international migration, and simultaneously achieves the Objectives 1 and 2.1. Specifically, this chapter discovers that there is a U-shaped selection pattern of migrants from inland to coastal regions utilizing four periods of Census data and a binary logit model. To explain this unique pattern, this chapter incorporates the *hukou* system into the self-selection framework, suggesting that the *hukou* system influences migrant selection through restricting *hukou*-related social amenities and distorting return to skills by wage discrimination. Then, the conditional logit model is further employed to explore these two possible impact channels. Empirical results verify the U-shaped migrant selection pattern on education and find that distorted return to education and *hukou*-related social amenities are two essential impact channels of the *hukou* system reshaping the migrant selection pattern.

The U-shaped selection pattern is different from the selection pattern of other internal or international migration cases. In the absence of migration restrictions, positive selection is prevalent, as is found in internal migration evidence. Inversely, international migrants tend to face direct skill-based restrictions posed by selective migration policies, making positive selection more common (Beine et al., 2011; Clark et al., 2007; Mayda, 2010). This research proposes another direction of how selective migration policies influence the migrant selection pattern through indirect channels. The *hukou* system does not directly prohibit the free movement of migrants but indirectly influences their economic returns and living costs in destination regions. Interestingly, this mechanism undermines mid-skilled migrants most instead of low-skilled ones, leading to the U-shaped selection pattern. The analysis of the *hukou* system and its mechanism will provide insights into how migration policies induce "segmented citizenship" and thus hinder the assimilation or integration of undocumented migrants in destination regions/countries.

These results are also inconsistent with the selection pattern of rural-urban migrants in China (Du et al., 2005; Luo & Xing, 2016; Wu, 2010; Xing, 2014). This difference may be induced by the different measures of skill levels. Yet, the more critical reason may be the difference in *hukou* transfer difficulties. Rural-urban migration within a region only involves the *hukou* status transfer (*nongzhuanfei*), while inter-regional migration involves the transfer of *hukou* status and location. Given that recent *hukou* reforms all concentrate on abolishing *hukou* status transfer, the *hukou* barriers are lowering for rural-urban migration within a region but remains high for inter-regional migration. Consequently, the degree of distortion by the *hukou* system is different, leading to different selection patterns. Furthermore, rural-urban migration and inter-regional migration are inherently intertwined. Rural-urban migrants may move across cities, while inter-regional migrants are mainly from rural areas, making the selection pattern more complex.

Although it has been going on since the reform and opening up in the 1980s, the *hukou* reform is still far from enough. Since 2014, China issued the "National New-type Urbanization Plan 2014-2020", which only relaxed the control over the urban *hukou* registration in small cities and towns, while strict restrictions still exist in big cities to limit population expansion (Wang et al., 2015; Zhang et al., 2019). Given that the migration inflow in large coastal cities is far more extensive than in other cities, the situation of most migrants has not improved, especially low- and mid-skilled migrants. Therefore, continued *hukou* reform and a unified national welfare system for migrants are required to reduce the difference in migration returns between different education groups and alleviate the distorted selection pattern.

CHAPTER 4. INCOME INEQUALITY, INEQUALITY OF OPPORTUNITY, AND MIGRANT SELECTION

4.1. Introduction

Chapters 4 and 5 will focus on the return to skills channel influencing migrant selection. This chapter will first discuss the effects of different income inequality components and achieve Objective 2.2. Based on the theoretical framework, it is hypothesized that lower levels of IOP and IOE in destinations increase migration incentives relatively more significantly for low-educated migrants. Employing the first wave of China Family Panel Studies (CFPS) survey data in 2010, province-level IOP and IOE are measured by the parametric method based on the ex-ante approach. Then, supported by the *2015 one percent population sampling survey* data, this research discusses the migrant selection in China and the effects of different inequality components.

The remainder of this chapter is organized as follows. Section 4.2 presents the literature review on the inequality of opportunity and develops a research framework to discuss its impact on migrant selection. The measure of inequality of opportunity is proposed in Section 4.3. Section 4.4 describes the empirical model design and data source used in the current study. Section 4.5 presents the empirical results, and Section 4.6 concludes this chapter.

4.2. Literature review and research framework

4.2.1. The inequality of opportunity

Traditional welfare egalitarianism and most economic analysis of inequality pay much attention to the equality of individual outcomes. Nonetheless, this thought is somewhat against people's intuition since it neglects the critical role of personal choices. Some political philosophers initially took this critique seriously and proposed that not all sources of inequality are seen as equally objectionable (Rawls, 1971; Sen, 1980). In their narrative, personal outcomes are determined by a combination of innate endowments that individuals are irresponsible for and enduring actions that individuals are responsible for. Rawls (1971) argues that the inequality induced by innate endowments is ethically unacceptable; society needs to offset inequality in outcomes attributable to innate endowments but not those components attributable to personal actions, which is just the concept of "equality of opportunity."

Previous theoretical and empirical evidence has identified the critical role of these innate endowments in personal outcomes. For example, people are born with gender and cannot change it. Nevertheless, this inherent endowment can significantly influence their life outcomes. Some scholars have found that gender discrimination will influence access to education and schooling outcomes, given the evidence that women behave commonly worse than men in education (Jacobs, 1996). Moreover, men are usually more advantageous in the labor market and earn higher wages even with similar education levels to women (Bobbitt-Zeher, 2007). The innate family background is also essential for personal growth and development, proved by the literature on intergenerational mobility. Chiu (1998) argues that children in wealthier families tend to have more access to training and education, which facilitates their economic performance in the future. These innate endowments are always unequal across different social strata, inducing a part of outcome inequality, i.e., inequality of opportunity.

From an economic perspective, Roemer (1998) proposed an algorithm for calculating IOP based on the above distinction with two new terms: circumstances and efforts. Circumstances are endowments beyond personal control and thus are exogenous to the person, such as gender, race, and family background. In contrast, efforts are decided by personal choices, such as occupation choice, learning/working time. Based on this distinction, one can divide individuals into several population groups according to their circumstance combinations. Within each group, individuals have the same circumstances but different degrees of effort. "Equality of opportunity" is realized when each individual has the same outcome levels if putting in the same effort levels, no matter which subgroup he/she belongs to. In such a case, one can regard the "inequality of opportunity (IOP)" as the between-group inequality of all groups and the "inequality of effort (IOE)" as the within-group inequality of all individuals.

Numerous circumstances can influence personal outcomes. Some discriminations among groups are the most common circumstances, such as race discrimination (Lang & Kahn-Lang Spitzer, 2020), gender discrimination (Bobbitt-Zeher, 2007), and others. Furthermore, scholars have also found evidence that social networks, education quality, and intergenerational inertia have substantial effects on personal outcomes (Arrow et al., 2018; Blume & Durlauf, 2000; Loury, 1989). These circumstances are largely associated with one's innate status and family support, such as the provision of social connections, skill training, genetic transmission, and the instilling of preferences and aspirations (Dardanoni et al., 2006). On the other hand, circumstances will also influence outcomes indirectly through their effects on personal efforts (Bourguignon et al., 2007). Put differently, personal effort is endogenous since circumstances can influence personal choices, such as education, employment decisions, and insertion in the labor market. Owing to data limitations, gender, race, parental educational level or occupation, and region of birth are the causal determinants most commonly used to measure circumstances. For example, intergenerational mobility, which measures the elasticity between paternal earnings and children's adult earnings, is the universal method in the literature to capture these circumstances' effects, given that parental features largely determine the circumstances mentioned above (Björklund et al., 2012; Fan et al., 2021; Sieg et al., 2020; Torche, 2015).

Thus far, empirical research into the consequences of IOP has been rare. Generally speaking, income inequality by unequal circumstances will damage the human capital accumulation because children in low-income families inherently face more liquidity constraints and other potential flaws for education, which undermines the economic performance in the long run. This mechanism is generally reflected in the literature on income inequality and intergenerational mobility as well as their consequences on economic performance (Chiu, 1998; Galor & Zeira, 1993; Mejía & St-Pierre, 2008; Sieg et al., 2020). In recent years, some scholars have begun to estimate the effects of IOP directly. Marrero and Rodríguez (2013) decomposed income inequality into IOP and IOE and investigated their effects on economic growth. The results show that the two kinds of inequality affect economic growth in opposite directions, while IOP generates negative effects. Song and Zhou (2019) and Song et al. (2020)developed a county-level index of IOP in China and investigated its impact on household behavior. The results show that IOP reduces household education expenditure but raises the probability and share of household risky asset investment. To the best of our knowledge, no research has paid attention to the role of IOP in population migration. Considering that migration will also influence human capital accumulation, it is interesting and crucial to investigate whether IOP will influence population migration and, thus, human capital accumulation, which is the research gap this chapter aims to fill.

4.2.2. Inequality of opportunity and migrant selection

Few studies have discussed differential roles of income inequality components in the educational selection of migrants, which may show a complex relationship. Generally speaking, IOE exerts direct effects on skill selection. It is intuitive and proven empirically that high-skilled migrants pursue high returns to skill, while low-skilled choose the opposite (Belot & Hatton, 2012; McKenzie & Rapoport, 2010). IOE is highly correlated with return to skills since personal efforts explain a large part of personal skill levels. As such, regions with a high (low) level of IOE will attract high(low)-skilled migrants to maximize their income level.

Inversely, IOP should have an indirect impact mechanism. On the one hand, IOP crowds out individuals with poor circumstances who tend to be low-skilled. Regarding gender, in a region with a high degree of IOP, women face severe gender discrimination in the labor market and earn lower wages than men with equivalent skill levels. In such a case, male individuals enjoy the benefits of inequality and prefer staying in this region. In contrast, female individuals suffer from severe wage loss and prefer moving to a region with a low degree of IOP. A large amount of literature has found that women occupied increasingly higher proportions of migrants (Even more than 50%) in recent decades, especially from those unequal regions/countries (Barber, 2000; Durand & Massey, 1992; Martin, 2007). Noting that women tend to have less education than men because of discrimination in education simultaneously, a high degree of IOP actually crowds out more low-skilled individuals. Regarding family backgrounds, similarly, individuals with better family backgrounds should prefer staying in regions with higher degrees of IOP since they can get higher wage levels relying on their family networks and other family benefits. In contrast, individuals with worse family backgrounds would like to escape this region. Given that individuals with worse family backgrounds are more likely to be low-skilled, IOP actually positively selects migrants regarding skill levels.

In summary, regions with high degrees of IOP will crowd out individuals with poor circumstances since they gain fewer outcomes than other individuals, even with the same level of effort. In the meantime, their poor circumstances will cause them to have less access to education/training and thus lower skill levels (Gamboa & Waltenberg, 2012; Palomino et al., 2019). As a result, more low-skilled migrants will migrate from regions with high degrees of IOP to regions with low degrees of IOP, that is, a positive selection of migrants.

On the other hand, migration is also beneficial for migrants' offspring when moving to more equal places. Ward (2020) found that internal migration facilitates intergenerational upward mobility among poor households, while this effect is not significant among wealthy households. If moving to regions with low IOP, children from poor households obtain more opportunities to accept better education and move upward along the income distribution. In contrast, a high degree of IOP will discourage poor households in children's human capital investment (Song & Zhou, 2019). Yet, this may not be the case in the context of China. Because of the household registration (*hukou*) system, low-skilled migrants do not have citizenship in the destination, and thus their children cannot enjoy equivalent education resources to natives. For example, they can only choose low-quality migrant schools after the compulsory education stage and cannot take the college entrance examination at the destination (Sieg et al., 2020). As a result, many children were left behind in their hometowns and received poor education and training (Lu, 2012). Golley and Kong (2013) proved that rural-urban migrants have high intergenerational mobility because their children are actually moving down the education ladder relative to themselves. Consequently, it is doubtful in China that migration benefits low-skilled migrants' offspring when moving to more equal places.

To verify above assumptions and arguments, this chapter proposes two hypotheses as shown below:

H4.1: Lower levels of inequality of effort in destination increases migration incentives more for low-educated migrants.

H4.2: Lower levels of absolute/relative inequality of opportunity in destination increases migration incentives more for low-educated migrants.

4.3. Inequality of opportunity calculation

4.3.1. The estimation method

This subsection first introduces a simplified conceptual framework of IOP to guide the calculation procedure, following Roemer (1998) and Ferreira and Gignoux (2011). Given a finite population of discrete individuals indexed by $i \in \{1, \cdots, N\}$, each individual i has a specific outcome y_i (i.e., income in this chapter) determined by his/her circumstance characteristics set C_i and effort level E_i, such that $y_i = f[C_i, E_i]$. As discussed by Roemer (1998), circumstance characteristics are personal attributes beyond individuals' control, such as gender, race, and family background, while the effort level represents the personal attributes under individuals' control, such as personal choice and studying/working time. Accordingly, the effort level E_i is taken as a continuous variable, while the circumstance set C_i is a vector of circumstance elements for individual i. Generally speaking, circumstances are exogenous given that they are beyond individuals' control, but personal effort levels can be influenced by circumstances, considering that circumstances will influence personal decision-making (Bourguignon et al., 2007).

To capture the IOP, one can separate all populations into K subgroups based on different circumstance combinations, which is given as $\Lambda \in \{T_1, \cdots, T_k\}$, such that $T_1 \cup T_2 \cup \cdots \cup T_k =$

$\{1, \cdots, N\}$, $T_l \cup T_k = \emptyset, \forall l, k$. In each subgroup type, individuals are homogeneous regarding circumstances, such that $C_i = C_j, \forall i, j | i \in T_k, j \in T_k, \forall k$. Nonetheless, their effort levels are heterogeneous and can be captured as $E_i = e^k(\pi)$, where $\pi \in [0,1]$ denotes the individual's quantile of effort distribution in subgroup T_k. Based on such a setting, Roemer (1998) proposed an ex-post approach that the equality of opportunity realizes when people obtain the same outcome if they exert the same degree of effort π across all subgroups, such that $y^k(\pi) = y^l(\pi), \forall \pi \in [0,1]; \forall T_k, T_l \in \Lambda$. In contrast, Van De Gaer (1993) proposed an ex-ante approach with a weaker criterion only requiring the same mean outcome for different population subgroups when realizing equality of opportunity, such that $\mu^k(y) = \mu^l(y), \forall l, k | T_k \in \Lambda, T_l \in \Lambda$, where $\mu^k(y)$ denotes the mean outcome of individuals in the subgroup T_k. Both approaches are relevant and plausible, but the latter one is more concise and easier to operate. Following the ex-ante approach, one can define the IOP as the between-group inequality, that is, $\theta_a^p = \mathrm{G}(\{\mu^k\})$ and $\theta_r^p = \mathrm{G}(\{\mu^k\})/\mathrm{G}(\{y_i\})$, where $\mathrm{G}(\ )$ denotes the inequality index algorithm; $\{\mu^k\}$ denotes the smoothed distribution of mean outcomes; $\{y_i\}$ denotes the actual income distribution; θ_a^p and θ_r^p represent absolute and relative IOP index, respectively.

According to this conceptual framework, one can measure the IOP parametrically or non-parametrically based on the ex-ante approach. The non-parametric method decomposes the total outcome inequality into between-group and within-group components, while the former is regarded as the IOP (Checchi & Peragine, 2010; Lefranc et al., 2008; Marrero & Rodríguez, 2013). However, this method tends to underestimate the IOP because circumstances may also influence the outcome through their effects on efforts. The parametric method alleviates this underestimation at the cost of some functional form assumptions (Ferreira & Gignoux, 2011; Song et al., 2020; Song & Zhou, 2019). As such, this chapter employs the ex-ante approach by parametric method following Ferreira and Gignoux (2011).

First, the determinants of income are divided into circumstances C and efforts E. As such, one can rewrite the general form of a stylized model of income $y = f[C, E, u]$ as:

$$y = f[C, E(C, v), u], \tag{4.1}$$

where v and u capture other stochastic errors influencing income. To measure IOP, one can have the reduced form of (1) as $y = \phi[C\psi, \varepsilon]$. This equation can be estimated by OLS based on its log-linearized version, $\ln y = C\psi + \varepsilon$. The coefficient ψ captures both the direct and indirect (through efforts) effects of circumstances on income y. Based on this estimation, one can estimate the predicted income contributed by circumstances:

$$\tilde{\mu}_i = \exp\left[C_i\hat{\psi}\right], \tag{4.2}$$

where $\tilde{\mu}$ is the counterfactual income level. Then one can estimate inequality degrees of income y and counterfactual income $\tilde{\mu}$ using an inequality index algorithm (G):

$$\theta^t = \mathrm{G}(y), \tag{4.3a}$$

$$\theta_a^p = \mathrm{G}(\tilde{\mu}), \tag{4.3b}$$

$$\theta_r^p = \theta_a^p / \theta^t \text{, and} \tag{4.3c}$$

$$\theta^e = \theta^t - \theta_a^p \tag{4.3d}$$

where, θ^t, θ_a^p, θ_r^p, and θ^e are total income inequality (TIE), absolute inequality of opportunity (IOP), relative inequality of opportunity (RIOP), and inequality of effort (IOE), respectively.

4.3.2. The data source and design

This chapter employs a representative survey database, the *China Family Panel Studies* (CFPS), to calculate the inequality indices. The CFPS, launched by Peking University, is a nearly nationwide, comprehensive, longitudinal social survey, providing rich information about income, occupation, and household (Xie & Hu, 2014). The current study employed the first wave of survey data in 2010. This wave covered 14798 households and 33600 individuals in 25 provincial administrative regions (excluding Hong Kong, Macao, Taiwan, Xinjiang, Tibet, Qinghai, Inner Mongolia, Ningxia, and Hainan). Xie and Hu (2014) compared the CFPS sample population structure with the *2010 National Census* data. They found a high degree of consistency, implying the high reliability of this database for calculating indices. The samples focused on individuals in the labor market aged between 18 and 65 years. To identify employment status, the CFPS records individuals' subjective responses to the following question: "do you currently have a job?" Following this, the samples removed those individuals without a job. Finally, a sample of 11,712 individuals for further calculation is obtained.

Firstly, the personal monthly income level in CFPS is used as the outcome. Before estimating the inequality index, it is needed to exclude the age effect influencing personal income. Conditional on other factors, individuals' age will influence their income level, caused by neither circumstances nor personal efforts. As such, this research restricts samples to individuals between 18 and 65, and regresses natural logarithm of income levels on age and age squared following Checchi and Peragine (2010). Then, this research takes the natural exponent of residuals (plus constant) after regression as the outcome into inequality estimation.

Secondly, this research discusses the choice of circumstance characteristics. As discussed in Section 2.1, multiple characteristics/elements determine personal circumstance combination C_i. The ideal way is to find all these characteristics to estimate the IOP. Given the availability of data sources, this research can only choose some most representative characteristics to capture circumstances. In general, gender and parental education level can essentially capture personal circumstances and are widely used in previous studies (Ferreira & Gignoux, 2011; Marrero & Rodríguez, 2013; Song et al., 2020; Song & Zhou, 2019). This research also introduces these variables into estimation: the dummy variable of gender (male = 1), maternal schooling years, and paternal schooling years. In addition, the *hukou* status (rural VS. urban) in China affects personal outcomes since childhood through its influence on educational resources, social environment, and so forth. Therefore, this research also takes the born *hukou* status of individuals inherited from their parents into consideration to measure IOP, that is, the dummy variable of 3-year-old *hukou* status (urban = 1) (Song et al., 2020; Song & Zhou, 2019).

Notably, since circumstances are still not exhausted, this method is a lower-bound estimation of overall IOP (Ferreira & Gignoux, 2011). Because this research only focuses on the effects of IOP on migrant selection, this issue will not affect the final result if the index calculation of each province uses a unified algorithm. Finally, this research employs the algorithm (G) of mean logarithmic deviation to measure inequality indices at the provincial level. Detailed estimation procedures and final calculated indices are presented in Appendix A1.

4.4. Empirical model design and variables

4.4.1. Empirical model design

This subsection presents the self-selection model used in this research. Migrant educational selection, the difference in migration probability between high- and low-educated individuals, can be directly represented by the log odds of emigration rates between high- and low-educated individuals (Belot & Hatton, 2012; Grogger & Hanson, 2011; McFadden, 1974). According to Borjas (1987), only income inequality and heterogeneous migration cost influence migration selection because other factors that do not vary with skills (such as average income level) have been offset (please see Appendix A2 for detailed model specification). Based on this specification, the empirical self-selection model is constructed:

$$\ln \frac{M_{ij}^{h}/N_{ii}^{h}}{M_{ij}^{l}/N_{ii}^{l}} = \alpha + \beta(\theta_j - \theta_i) + \boldsymbol{X_{ij}\gamma} + \delta_i + \sigma_j + \varepsilon_{ij}, \tag{4.4}$$

where $M_{ij}^{h(l)}$ represents the number of high(low)-educated migrants from region i to region j;

$N_{ii}^{h(l)}$ represents the number of high(low)-educated natives staying at region i; $\theta_{j(i)}$ represents the inequality indices at region $j(i)$, including TIE, IOP, RIOP, and IOE; $\boldsymbol{X}_{ij}$ is a vector of control variables capturing heterogeneous migration costs (discussed in section 4.4.3); α, β, and $\boldsymbol{\gamma}$ are coefficients that aim to estimate; δ_i and σ_j control the fixed effects of origin regions and destination regions.[5] This empirical model is mainly concerned with the significance and values of β. $\beta > 0$ implies that regions with a high(low) level of inequality increase migration incentives more for high(low)-educated migrants (i.e., positive selection), and the opposite is negative selection.

Furthermore, it is noticed that previous migration flows (and self-selection) may influence current income structures in destination regions and thus their inequality levels, which, in turn, influence the current migration flows (and self-selection). To eliminate this concern, three measures are introduced to ensure the robustness of results: First, the sample removes all past migrants when calculating inequality indices to do a robustness test. Excluding past migrants when calculating inequality indices ensures that the component of inequality influencing migration selection is induced by inherent natives instead of past migrants. The CFPS provides detailed information about the place of birth and current residence, which enables us to separate past migrants whose current residence is different from their places of birth. Secondly, the

5 All provinces are divided into seven regions, including the North China region (Beijing, Tianjin, Hebei, Shanxi, Inner Mongolia), Northeast region (Liaoning, Jilin, Heilongjiang), East China region (Shanghai, Jiangsu, Zhejiang, Anhui, Fujian, Jiangxi, Shandong), Central China region (Henan, Hubei, Hunan), South China region (Guangdong, Guangxi), Southwest region (Chongqing, Sichuan, Guizhou, Yunnan), and Northwest region (Gansu).

empirical model introduces the provincial migrant share of all people into the independent variables to control the effects of past migrant stock. Thirdly, this research calculates all inequality indices in 2010 before migration to guarantee that current migration will not influence the inequality. Notably, these measures can only alleviate this endogenous issue to a large extent but not completely solve it, implying that the current analysis is not a rigorous causal analysis. However, one can still obtain sufficient information and implications from empirical results.

4.4.2. Migration flows

This research employs micro-level data from the 10% subsample of the *2015 one percent population sampling survey*, implemented by the China National Bureau of Statistics (NBS), to capture bilateral migration flows at the provincial level. This database includes 1,371,252 individuals representing 1‰ of China's total population. As this chapter focuses on economic migration, only the migration of individuals in the labor market is considered. Since the survey reports each individual's employment status, one is able to directly separate out individuals in the labor market. In addition, the sample eliminates all individuals below 18 and above 65 years old.

To investigate the difference in migration among different education levels, all individuals are divided into two skill types: i) high-educated individuals with a college diploma or above and ii) low-educated individuals with a senior high school diploma or below. The final sample was comprised of 659,479 individuals, of whom high-educated individuals and low-educated individuals accounted for 14.5% and 85.5%, respectively. According to the survey design, migrants are defined as individuals whose current place of residence is different from their residence five years ago at the province level. In other words, the survey identifies interprovincial migrants who migrate in five years. After sorting, 37,164 migrants were identified who had undergone interprovincial migration between 2011 and 2015. Based on these individual-level data, one can calculate the bilateral migration flows by skill types across 31 provincial administrative regions (excluding Taiwan, Hongkong, and Macao). Then, this research further calculates the emigration rate (M_{ij}^{s}/N_{ii}^{s}) at the region-pair level by the ratio of the number of high (low)-educated migrants between two provinces to the number of high (low)-educated stayers at the original province. The descriptive statistics of these individuals are shown in Table 4.1, and aggregated emigration rate data is shown in Appendix A3.

Table 4.1 Descriptive statistics of individuals

	All individuals		Migrants		High-educated migrants		Low-educated migrants	
	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.
Age	40.160	11.599	33.036	10.067	28.414	6.346	33.763	10.347
Male	0.573	0.495	0.635	0.481	0.589	0.492	0.642	0.479
Unmarried	0.145	0.352	0.296	0.457	0.519	0.500	0.261	0.439
Years of schooling	9.691	3.376	10.028	2.984	15.707	1.021	9.135	2.067
Observations	659,479		37,164		5,050		32,114	

4.4.3. Heterogeneous migration costs

This section will discuss variables capturing heterogeneous migration costs. Firstly, in terms of the migration cost induced by distance, high-skilled migrants are more likely to obtain new information and learn faster, and thus their migration costs are less. In contrast, such costs for low-skilled migrants are relatively higher and restrict their migration. Since this migration cost is highly associated with physical distance, the Euclidean distance between provincial capitals (*Distance*) is used to measure migration cost.

Secondly, migrants with different education levels face different labor market conditions, while high-educated migrants tend to prefer regions with more population, skill-intensive industries, and job opportunities. Therefore, four variables are used to capture labor market conditions. This research first employs the permanent *population* at the end of the year by provinces in 2010 to capture the population agglomeration effect. Then, this research uses the urban *Unemployment rate*, *Industry structure* (the GDP percentage ratio of the tertiary industry to the secondary industry), and *Foreign direct investment* amount to capture the regional difference in the labor market. All data is collected from the China Statistical Yearbook in 2010.

Thirdly, previous scholars have shown that migration networks significantly affect migrant selection (McKenzie & Rapoport, 2010). According to their theoretical and empirical study, migrant networks are more beneficial for low-skilled migrants than high-skilled migrants. This variable will also alleviate the endogenous issue, as discussed in Section 4.1. Based on the *Chinese 2010 Census microdata*, this research calculates the existing migrant stock (*Migration stock ratio*) between each pair of origin and destination locations in 2010.

Finally, migration policy also plays a significant role in migrant selection (Belot & Hatton, 2012). Although there is no strict selective immigration policy between provinces, the unique household registration (*hukou*) system still obstructs the free migration of China's population. This system does not forcibly hinder people's physical movement. Instead, it considerably affects their living standards after migration because the local *hukou* determines the amenities they have access to in the destination, which involves medical care, children's education, housing, and social welfare (Hui et al., 2014; Tao et al., 2015). Therefore, the different ability/possibility to obtain local *hukou* largely influences individuals' migration decisions. Unfortunately, this ability/possibility is hard to measure directly since different cities have different *hukou* policies. The optimum way is to collect and summarize government documents (Kinnan et al., 2018; Zhang et al., 2019). However, on the one hand, these files are difficult to obtain, and on the other hand, it is also difficult to quantify different files. An alternative way is to use the ratio of all populations to the *hukou* population by birth to proxy the degree of *hukou* restriction (Cai et al., 2001). This research follows the second method and attempt to proxy the *hukou* restriction by the ratio of migrants who have not obtained local *hukou* during 2011~2015 to all migrants (*hukou restriction*), based on *2015 one percent population sampling survey* data. In this case, the larger the *hukou restriction*, the lower the chance of obtaining local *hukou*, and the more restrictive the *hukou* policy.

4.5. Empirical results

4.5.1. Descriptive results

This subsection first plots the distribution of inequality indices and aggregated emigration rates across provinces in Figure 4.1 and Figure 4.2. Regarding TIE and IOE, according to Figure 4.1, there is an increasing tendency from coastal provinces to inland provinces (including central and western provinces). In other words, western provinces are relatively unequal regions, while coastal provinces are relatively equal regions. Regarding IOP, the distribution is relatively more complex, displaying that the most equal and unequal provinces coexist in inland regions, while inequality degrees in coastal regions are moderate. Figure 4.2 shows that provinces near coastal regions suffered the most significant population loss, including both high- and low-educated individuals, as they are spatially close to the epicenter of China's economic development, thus incurring lower migration costs. Secondly, there are varying degrees of population loss in provinces in the northeast and west, such as Heilongjiang, Sichuan, and Chongqing. Finally, the emigrant selection pattern, the difference in the emigration rate between high- and low-educated migrants, varies with regions. Western provinces (e.g., Guizhou, Sichuan, Yunnan) and central provinces (e.g., Henan, Anhui) experience a significantly negative emigrant selection, in that the emigration rate of low-educated migrants surpasses that of high-educated migrants to a great extent. In contrast, the coastal provinces (Jiangsu, Zhejiang) and northeast provinces (Heilongjiang, Liaoning) show a significantly positive selection of emigrants.

To further demonstrate the relationship between inequality and emigration rate, Figure 4.3 draws the scatter plot between aggregated emigration rate and four inequality indices. There is a clear negative correlation between aggregated emigration rate and total income inequality (TIE), implying that more low-educated migrants emigrate from regions with a higher level of income inequality, consistent with the self-selection framework. Considering that internal migration in China is mainly from the underdeveloped central and western regions to the developed coastal regions (Shen & Liu, 2016), this relationship intuitively validates the hypothesis that the relatively lower degrees of income inequality (TIE) in the coastal areas lead to higher migration probabilities of low-educated ones in China's internal migration. Similarly, there is also a negative correlation between emigration rate and inequality of opportunity (IOP, RIOP) and effort (IOE), implying that more low-educated migrants emigrate from regions with a higher level of IOP and IOE. These descriptive results initially verify proposed two hypotheses.

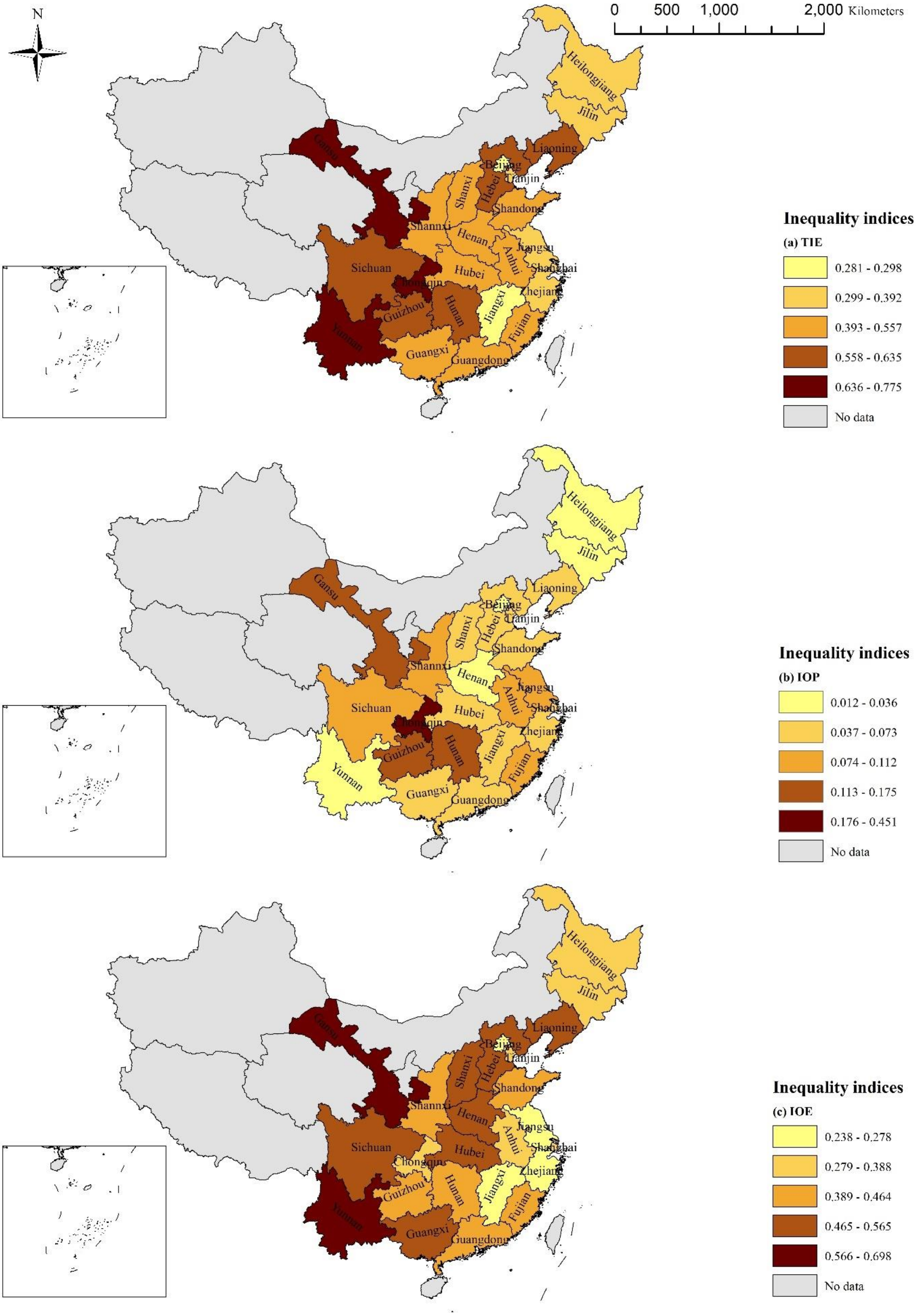

Figure 4.1 The distribution of inequality index

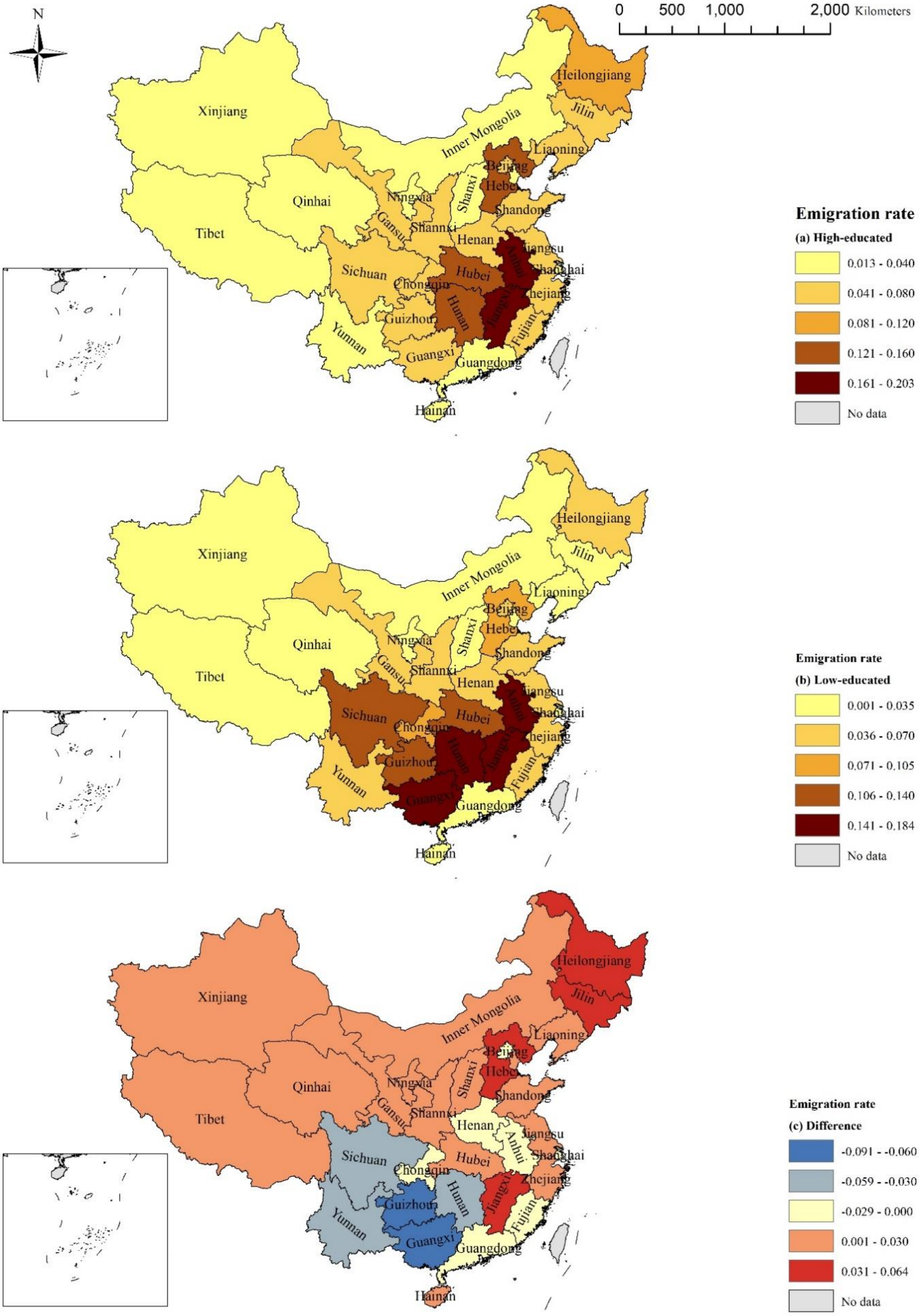

Figure 4.2 Aggregated emigration rate in China by provinces.

Note: the aggregated emigration rate is calculated as $\sum M_{ij}^{H(L)}/N_{ii}^{H(L)}$.

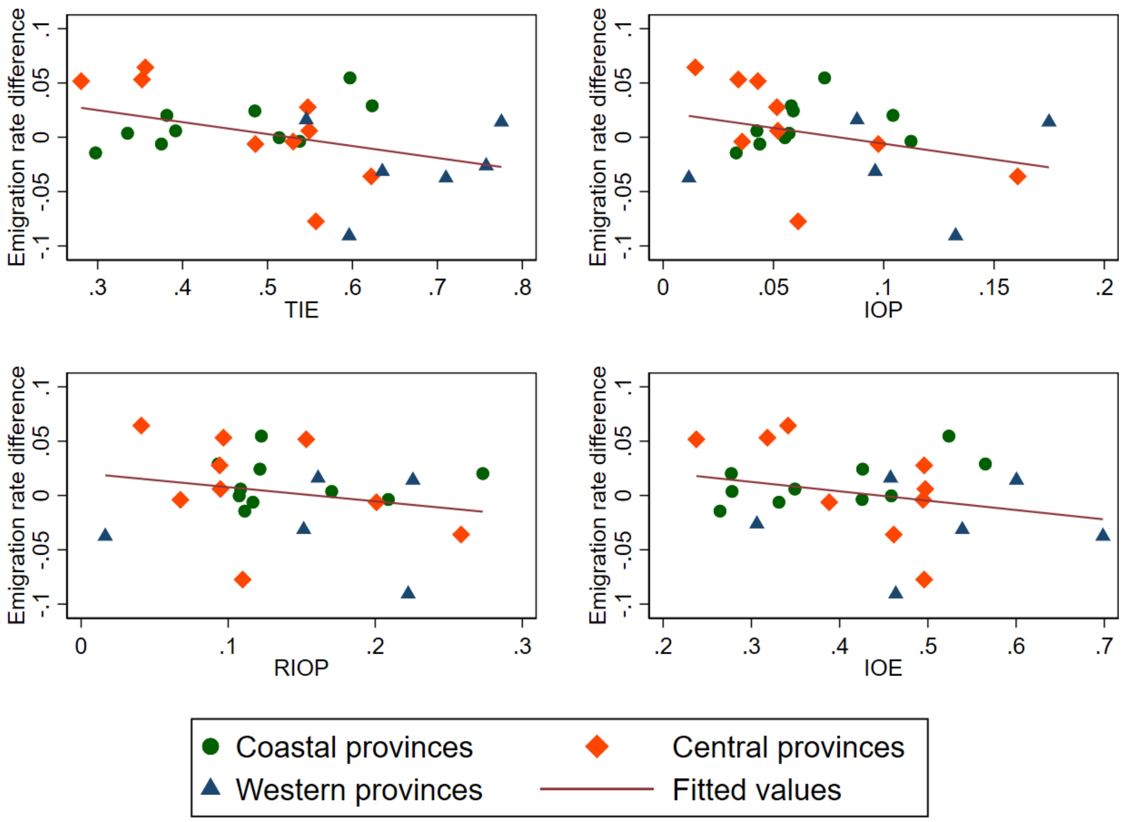

Figure 4.3 Inequality indices and aggregated emigrate rate difference in China.

Note: Coastal provinces includes Beijing, Tianjin, Hebei, Liaoning, Shanghai, Jiangsu, Zhejiang, Fujian, Shandong, Guangdong, Hainan; Central provinces includes Shanxi, Jilin, Heilongjiang, Anhui, Jiangxi, Henan, Hubei, Hunan, Guangxi; Western provinces includes Inner Mongolia, Chongqing, Sichuan, Guizhou, Yunnan, Tibet, Shaanxi, Gansu, Qinghai, Ningxia, Xinjiang.

4.5.2. Regression results

Table 4.2 shows the regression results of the empirical selection model. Since the CFPS database only covers 25 provincial administrative regions for the inequality indices calculation, and some origin-destination pairs do not have migrants, this research only includes 425 observations in the regression. All models pass the F statistic test, and the adjusted R^2 reaches 0.3 after including controlling variables in Columns (5) to (9) of Table 4.2. Regarding key coefficients, Columns (1) and (5) of Table 4.2 shows that the total inequality (TIE) have significant positive effects on migrant selection. This implies that high/low-educated migrants tend to migrate to regions with high/low degrees of income inequality to maximize their income levels, consistent with Borjas's initial hypothesis. Given that coastal provinces have relatively lower levels of total inequality (TIE), this result explains why such an enormous number of low-educated migrants have moved from inland provinces to coastal provinces over the past decades.

Similarly, Columns (4), (8), and (9) of Table 4.2 confirm the significant positive effects of IOE on migrant selection. In other words, equality of effort in destination increases migration incentives more for low-educated migrants, which verifies H4.1. Furthermore, Columns (6), (7), and (9) of Table 4.2 also shows that IOP and RIOP exert significant positive effects on migrant selection after controlling other influencing variables. This result implies that low-educated migrants will also escape from those regions with high IOP levels and migrate to a region with more equal opportunities, even when controlling for IOE, which verifies H4.2. Finally, Column (9) of Table 4.2 indicates that low-educated migrants even pay more attention to the low level of IOP than IOE, given the slightly larger coefficient value of IOP. To ensure the robustness of results, this research additionally tests the adjusted inequality indices excluding past migrants when calculating. As shown in Column (10) of Table 4.2, two key inequality indices still have significant positive effects on emigration rate difference, confirming results' reliability.

These results demonstrate the phenomenon that low-educated migrants will be self-selected into regions with low levels of IOP and IOE. On the one hand, equality of effort means that low-educated individuals can get relatively higher wage levels compared with regions with high degrees of IOE in terms of equivalent education level, thus attracting more low-educated migrants. On the other hand, equality of opportunity means that income levels are explained more by personal efforts than circumstance factors. Given that low-educated migrants tend to have fewer circumstances advantages, they can rely more on personal efforts and obtain relatively more opportunities to get higher wage levels in regions with low degrees of IOP. This explains why low-educated migrants tend to be self-selected into regions with low levels of IOP and RIOP, even conditional on IOE.

Regarding other controlling variables, most of the results are in line with expectations. Firstly, the difference in population between destination and origin exert a significant positive effect on migrant selection. In other words, more high-educated individuals emigrate from less populated areas to densely populated areas than low-educated individuals. In contrast, a high unemployment rate discourages more high-educated migrants and thus leads to a negative selection of migrants. In addition, industry structure fails to explain the migrant selection well, while foreign direct investment leads to a negative selection of migrants. Secondly, the distance has significantly positive effects on migrant selection; that is, a longer migration distance will attract more high-educated migrants than low-educated migrants. In contrast, the migration stock ratio causes significant negative effects on migrant selection. In other words, a large migrant stock in the destination is more beneficial for low-educated migrants and thus will induce negative selection. Thirdly, the *hukou* restriction captured by the ratio of non-*hukou* migrants in the destination exerts insignificant effects on migrant selection, inconsistent with previous studies. The accuracy of the *hukou* restriction degree proxy may be responsible for this result because it is hard to measure *hukou* obtainment ability directly as different cities have different *hukou* policies.

Table 4.2 Regression results of selection model

	Dependent variable: emigration rate difference ($\ln\left[\left(M_{ij}^{H}/N_{ii}^{H}\right)/\left(M_{ij}^{L}/N_{ii}^{L}\right)\right]$)									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
OD Total inequality (TIE)	2.077***				1.351***					
	(0.375)				(0.349)					
OD Inequality of opportunity (IOP)		0.171				1.446***			1.681***	1.962**
		(0.487)				(0.486)			(0.491)	(0.912)
OD Relative inequality of opportunity (RIOP)			-0.167				0.918**			
			(0.68)				(0.381)			
OD Inequality of effort (IOE)				2.015***				0.852**	1.118***	0.898*
				(0.381)				(0.425)	(0.426)	(0.458)
OD log population					0.685***	0.754***	0.751***	0.726***	0.693***	0.668***
					(0.108)	(0.107)	(0.108)	(0.110)	(0.109)	(0.113)
OD unemployment rate					-0.011**	-0.010*	-0.010**	-0.014***	-0.010*	-0.010*
					(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
OD industry structure					-0.053	-0.082	-0.095	-0.114	-0.046	-0.058
					(0.073)	(0.073)	(0.073)	(0.072)	(0.073)	(0.077)
OD log foreign direct investment					-0.125***	-0.184***	-0.195***	-0.132***	-0.138***	-0.134***
					(0.041)	(0.040)	(0.042)	(0.044)	(0.044)	(0.042)
Des Hukou restriction					-1.110	1.021	0.902	-2.375	-0.278	-0.263
					(1.694)	(1.852)	(1.908)	(1.827)	(1.904)	(2.004)
Des migration stock ratio					-0.964*	-0.983*	-0.927*	-0.883	-0.987*	-0.987*

					(0.543)	(0.548)	(0.549)	(0.550)	(0.544)	(0.548)
Log distance					0.255***	0.238***	0.231***	0.251***	0.251***	0.244***
					(0.082)	(0.082)	(0.083)	(0.083)	(0.082)	(0.083)
Constant	0.298*	0.249	0.248	0.295*	-0.371	-2.242	-2.095	0.799	-1.115	-1.083
	(0.157)	(0.163)	(0.163)	(0.158)	(1.664)	(1.798)	(1.840)	(1.772)	(1.836)	(1.909)
Origin FE	Yes	Yes	Yes	Yes	No	No	No	No	No	No
Destination FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	425	425	425	425	425	425	425	425	425	425
Adjusted R^2	0.339	0.289	0.290	0.335	0.361	0.352	0.347	0.344	0.361	0.352

Notes: 1. OD means the variable difference between destination and original regions, while Des indicates the variable value of destination regions; 2. the calculation of inequality indices used in Column (10) only consider natives; 3. Robust standard errors are presented in parentheses; 4. ***, **, and * denote significance at the 1, 5, and 10 percent levels, respectively.

4.5.3. Heterogeneity

This subsection further investigates the heterogeneity between different migrant groups. Firstly, different age groups may have different migration incentives; thus, this research compares the migrant selection of different age groups, as shown in columns (1) to (4) of Table 4.3. The results demonstrate that inequality indices only significantly influence the self-selection of the young cohort instead of the old counterpart. Considering that young cohorts are the main force of migrants, the lack of observations of the old cohort may result in insignificance (Plane, 1993). However, the more important reason may be that young individuals are more sensitive to IOP levels. On the one hand, young individuals are at the beginning of their careers and thus are more likely to chase higher income levels and more social opportunities. On the other hand, they begin to consider the growth environment of the next generation, making them more concerned about the degree of IOP in society. As a result, income inequality and its two components exert more significant effects on the young cohort.

Secondly, this subsection also compares the difference in gender because different genders have different conditions in the labor market. The results in Columns (5) to (8) of Table 4.3 show that migrants of both genders are concerned about TIE, IOE, and IOP, while the coefficients of female migration flows are slightly smaller and less significant. This result implies that men and women are nearly equivalently sensitive to inequality. The possible reason is that men are usually the household head and main labor source in a family, they tend to be more sensitive to income inequality and its components. On the other hand, women are born with poorer circumstances and faces severe gender discrimination in the labor market, thus making them also sensitive to inequality. These two effects make men and women equally sensitive to inequality, and the former effect is more obvious.

Finally, this subsection compares the heterogeneity of destinations between coastal provinces and inland provinces, as shown in columns (9) to (12) of Table 4.3. The result demonstrates that TIE, IOE, and IOP all exert more significant effects when migrants' destinations are coastal provinces. In contrast, the coefficients become insignificant when destinations are inland provinces, as shown in Columns (11) and (12) of Table 4.3. As mentioned previously, the migration from the inland provinces to coastal provinces is the dominant migration pattern in China, which leads to the insignificance of the latter results.

Table 4.3 Heterogeneity of migrant groups

	Young cohort (Age 18~35)		Old cohort (Age 36~65)		Male		Female		Destination: Coastal provinces		Destination: Inland provinces	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
OD Total inequality (TIE)	1.634***		0.092		1.539***		1.238***		2.724***		0.639	
	(0.356)		(0.603)		(0.372)		(0.411)		(0.510)		(0.511)	
OD Inequality of opportunity (IOP)		1.392***		1.000		1.886***		1.439**		2.795***		1.069
		(0.506)		(0.868)		(0.523)		(0.597)		(0.740)		(0.687)
OD Inequality of effort (IOE)		1.803***		-0.431		1.294***		1.105**		2.684***		0.220
		(0.436)		(0.701)		(0.453)		(0.502)		(0.591)		(0.679)
Other controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Destination FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	400	400	163	163	376	376	299	299	206	206	219	219
Adjusted R^2	0.427	0.426	0.553	0.557	0.407	0.407	0.412	0.411	0.508	0.505	0.113	0.112

Notes: 1. OD means the variable difference between origin and destination regions, while Des indicates the variable value of destination regions; 2. Robust standard errors are presented in parentheses; 3. ***, **, and * denote significance at the 1, 5, and 10 percent levels, respectively.

4.6. Conclusions and discussions

This chapter investigates the role of IOP in the educational selection of migrants in China, which proposes another channel of how IOP influences regional human capital accumulation, achieving Objective 2.2. This research hypothesizes that lower levels of IOP and IOE in destination both increase migration incentives more for low-educated migrants than high-educated counterparts. Employing the CFPS database, income inequality, IOP, and IOE across 25 provinces are quantified. Then, this research estimates inequality indices' effects on migrant educational selection empirically based on interprovincial migration flow data calculated from the *2015 one percent population sampling survey* data. Finally, this research investigates the heterogeneous effects of different migrant groups.

Findings support the proposed hypotheses. Firstly, the internal migration in China shows that lower-educated ones have higher migration probabilities from inland provinces to coastal provinces. Two reasons contribute to this selection pattern: i) coastal provinces have relatively low levels of total inequality (TIE); ii) low-educated migrants prefer destinations with a low level of income inequality to compensate for their income. Secondly, results demonstrate that the IOE and IOP exert significant positive effects on migrant selection, implying that lower levels of IOP and IOE in destination indeed increase migration incentives more for low-educated migrants than high-educated migrants. Finally, results also show that young migrants who migrate to coastal regions pay more attention to inequality than other migrant groups.

This chapter demonstrates that IOP will influence regional human capital accumulation through population migration, contributing to the scant body of empirical research on the IOP. A high level of IOP will retain high-educated individuals and mitigate the brain drain, although it damages initial human capital formation. However, this consequence may be a misallocation of human capital. Since they face severe competition in coastal regions without equivalent social opportunities as in original regions, high-educated individuals would instead choose jobs that do not match their education levels in unequal but underdeveloped original regions. In other words, the IOP has distorted labor allocation across regions that should be regulated by the regional return to skills. This kind of labor misallocation may result in the waste of human capital and hence the loss of economic growth at the national level.

On the other hand, this chapter identifies the higher migration probabilities of low-educated ones in China induced by the regional disparity in different components of income inequality, entirely different from the typical positive selection of internal migrants in other countries (Bacolod et al., 2021; De la Roca, 2017). The past few decades have witnessed an enormous number of migrant workers migrating from inland rural areas to coastal megacities to escape poverty and make a living. The findings in this chapter indicate that these low-educated migrants are not only fleeing poverty but also chasing more social opportunities, which, to a large extent, determines both their own and their next-generation outcomes. However, given the growing urban problems due to overpopulation, these destination megacities have gradually introduced various policies to squeeze out low-educated/income migrants in recent years, which has largely impeded their possibilities to escape poverty and raise their living standards.

From a more macro perspective, instead of narrowing the regional development gap, this tendency will further enlarge regional disparity and exacerbate income inequality in both origin and destination.

CHAPTER 5. LABOR MARKET EFFECTS OF TECHNOLOGICAL CHANGE ON MIGRANTS: TAKING INDUSTRIAL ROBOTS AS A CASE STUDY

5.1. Introduction

This chapter still focuses on the income channel on migrant selection. Unlike chapter 4, this chapter turns to discuss the supply and demand shock of the labor market induced by technological change that also significantly influences wage return to skills, which achieves Objective 2.3. Yet, technological change is difficult to measure, given the difficulty of data collection. Fortunately, multipurpose industrial robots, that has been widely promoted and installed in China over the last decade have sufficiently detailed data and provide us with a small entry point to discuss labor market consequences of technological change.

Following Acemoglu and Restrepo (2020) and Dauth et al. (2021), to analyze the effects of exposure to robots on migrants, this chapter first utilizes the installation of industrial robot data from the International Federation of Robotics (IFR) to build a Bartik-type index to indicate exposure to robots at the prefecture city-level. Then, this chapter employs the migrant data from China Migrant Dynamic Survey to investigate how exposure to robots influences the occupation and skill selection of migrants. Results demonstrate that instead of replacing workers, exposure to robots booms the manufacturing industries, thus attracting more migrants to production occupations. However, the migrants attracted are mainly high-skilled (especially those with three-year college degrees), implying that industrial robots have replaced low-skilled migrants but attracted more high-skilled migrants to complement their tasks. The resulting positive selection of migrants exists within production and service occupations and across all occupations. This chapter further introduces an instrumental variable (IV) to solve the endogenous problem, and the results are still highly stable.

The rest of this chapter is as follows. Section 5.2 gives a research background of industrial robots in China. Section 5.3 reviews recent studies related to this research and then proposes several hypotheses. Section 5.4 introduces the empirical design and data sources. The empirical results are discussed in Section 5.5. Finally, Section 5.6 concludes and proposes areas for future studies.

5.2. Research background

Will machines completely replace labor in the future? This is a historical concern that can be traced back to the first industrial revolution. With the advent of the information age in the second half of the last century, it has captured people's attention and raised plenty of discussions and investigations in developed countries (Autor et al., 1998; Murphy et al., 1998). According to Frey and Osborne (2017), automation by machines jeopardizes about 47% of employment in the United States (the US). However, manufacturing industries are already heavily automated. The once imagined robots crept into the production lines from the laboratory to complete complex but precise tasks, such as painting, assembling, and even performing surgery, with very little assistance from humans. In addition to robots performing repetitive routines, researchers and high-tech companies are exploring ways to add more automation to other tasks not yet automated. Undoubtedly, this has intensified people's fears about whether these robots will replace workers and cause severe unemployment, especially in the manufacturing sector (Crowley et al., 2021; Leigh & Kraft, 2018). Thus, it is critical to answer the following question: "How do these robots influence the labor market?"

Almost all studies related to this question focus on developed countries, such as the US (Acemoglu & Restrepo, 2020; Graetz & Michaels, 2018) and Germany (Dauth et al., 2021), neglecting emerging economies. China is the most representative developing country undertaking fundamental economic transition and technological upgrading. The past three decades have witnessed China's economic miracle fueled by the unprecedented migration of labor from the unproductive farm sector to work in cities. Along with the new era of the knowledge economy and the Lewis turning point, the high productivity resulting from the once infinite pool of surplus rural labor slows down when the labor pool shrinks and labor costs rise rapidly, which requires firms to find new ways to replace the increasingly expensive labor, where the use of industrial robots is the effective choice (Cheng et al., 2019). This new track has also received significant attention and support from the government, as they utilize financial subsidies and other policies to promote the use of robots. As a result, robot adoption in China has maintained an annual growth rate of more than 30% since the early 2000s, as shown in Figure 5.1. In 2016, China became the world's largest user of industrial robots, with nearly 350,000 units of industrial robots in use. As demonstrated by the five-year plan of The Ministry of Industry and Information Technology, China aims to double the density of manufacturing robots from 2021 to 2025[6].

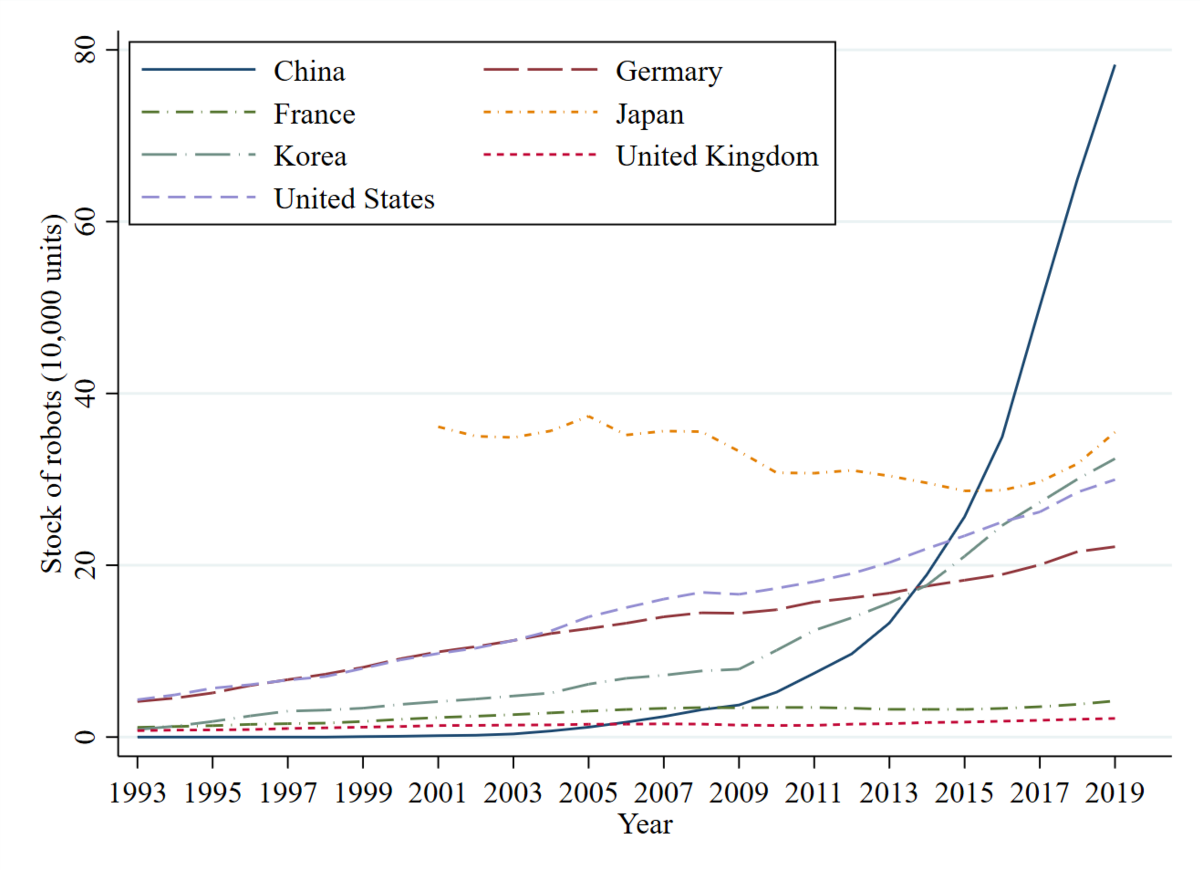

Figure 5.1 The stock of industrial robots in China and other developed countries from 1993 to 2019

Note: The data comes from the International Federation of Robotics (IFR).

Intuitively, such dramatic robot adoption should have profoundly impacted China's labor market. According to the report by Li et al. (2019), although robot adoption has improved

[6] Please see the CHINA DAILY news (*https://www.chinadailyhk.com/article/253763#China-aims-to-be-hub-of-global-robotics-industry*).

production efficiency by more than three times, machines have supplanted only 10% of humans in the Pearl River Delta, China. Surprisingly, only a few studies have investigated this crucial issue. This research aims to provide more evidence on the heated discussion about how robot adoption influences the labor market in China from the perspective of migrant workers. Migrant workers are the main source of labor in China (Freeman, 2015). However, unlike other countries, migrant workers in China are exceptional because of the *hukou* barriers that restrict them from social benefits, and they face severe labor market discrimination in their destination cities (as discussed in Chapter 3). Moreover, migrant workers tend to be low-skilled; over 50% have only junior high school education or below, as shown in Figure 5.2. In such cases, migrant workers tend to be more vulnerable to labor market shock and respond to industrial robot exposure more elastically. Therefore, this research analyzes migrants separately; this perspective is different from that of the existing literature.

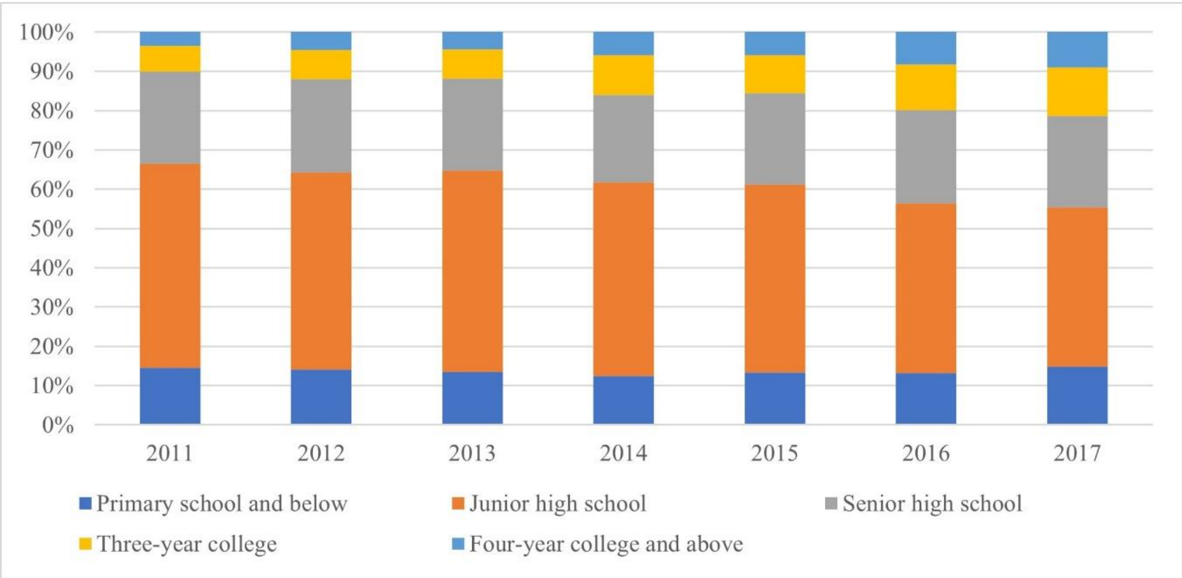

Figure 5.2 Education level distribution of migrants

Note: The data comes from China Migrant Dynamic Survey (CMDS).

5.3. Literature review and research framework

5.3.1. Technological change and labor market

The fear of unemployment induced by technology upgrading is not new; it can be traced back to the first industrial revolution in the early 19th century. In the classical economic discussion, technological changes are assumed to be nonneutral and skill-biased, thus leading to disproportionately higher demand for skilled workers and the skill premium of wage in the labor market (Katz & Autor, 1999; Katz & Murphy, 1992; Murphy et al., 1998). However, this argument fails to explain "skill polarization," which has been commonly observed in developed countries in the past decades, a phenomenon in which high- and low-skilled employment shares increase disproportionately than mid-skilled employment shares (Autor & Dorn, 2013; Gallie, 1991; Goos & Manning, 2007). To address this issue, Autor et al. (2003) proposed another approach to explain how technological change influences the labor market. Their seminal paper provided a new perspective to discuss labor replacement induced by technological change.

The basic idea in the study of Autor et al. (2003) is that from a "machine's-eye" view, each job comprises two tasks—routine and non-routine tasks. The former type of tasks involves repeated work with explicitly programmed rules, such as assembling, packaging, monitoring, and simply moving. These tasks can be simply codified into a computer program and, as such, are more easily to be substituted by automated machines. Interestingly, these tasks are often undertaken by mid-skilled (instead of low-skilled) labor, causing them to be more vulnerable to automation. By contrast, there are still many non-routine tasks that explicit "rules" or procedures cannot be drawn by computer code to accomplish them. Autor et al. (2003) further classified these non-routine tasks into abstract and manual tasks. Abstract tasks require cognitive skills, creativity, problem-solving, and complex communications capabilities, which are usually undertaken by high-skilled workers with a college education; manual tasks require visual and motor processing capabilities and are usually undertaken by low-skilled, low-educated workers.

In summary, technological changes will essentially substitute routine tasks but have no significant impact on non-routine tasks. Therefore, mid-skilled labor is most susceptible to technological advancement because of their routine task-intensive occupations. While revealing how machines replace workers, the task-based approach answers why "skill polarization" has appeared in developed countries in recent decades, which has been confirmed many reliable studies. Goos and Manning (2007) provided empirical evidence that the task-based approach can better explain the job polarization and increasing wage inequality in the United Kingdom since 1975. de Vries et al. (2020) investigated the effects of industrial robot adoption on jobs, which demonstrates that robot adoption leads to considerable loss of employment in routine-task jobs.

Technological change does not only substitute labor (i.e., *displacement effect*) but also creates new tasks and jobs where humans still have comparative advantages (i.e., *reinstatement effect*). The current automation is still in progress, implying that automated machines still need complementary humans to maintain, codify, and monitor them, which creates new jobs (Acemoglu & Restrepo, 2019). These new tasks generally require cognitive skills and high education level. Consequently, automation tends to create more jobs for high-skilled labor. Moreover, the progress of automation facilitates the productivity level of the economy, which then stimulates labor demand. Acemoglu and Restrepo (2018) proposed that productivity gain from automation creates more profit for firms and thus allows the economy to expand, which increases labor demand.

Because of these two confounding effects, the impact of automation on aggregate employment is quite uncertain, resulting in mixed empirical results in recent years (mainly focusing on industrial robots), including employment augmenting (Autor & Salomons, 2018), employment neutral (Dauth et al., 2021; Graetz & Michaels, 2018), and employment reduction (Acemoglu & Restrepo, 2020; Borjas & Freeman, 2019; de Vries et al., 2020). However, although the results of the studies of Dauth et al. (2021) and Graetz and Michaels (2018) show no significant effects of industrial robots on aggregate employment, they found a huge impact of industrial robots on low-skilled labor. Autor and Dorn (2013) proposed that local labor markets that have adopted information technology have reallocated the replaced low-skilled labor into services occupations, which was later empirically identified by Autor & Salomons (2018). These pieces of evidence support the thought that technological changes will increase the employment share of high-skilled labor and supplant low- and mid-skilled labors labors into the service sector simultaneously. As stated by Autor (2015), “Even if automation does not reduce the quantity of jobs, it may greatly affect the qualities of jobs available.”

Surprisingly, we know little about the condition in developing countries, although they are constantly catching up with the pace of developed countries in terms of technology. Because of the data availability in industrial robots, there is some empirical evidence about China. Following the robot exposure index proposed by Acemoglu and Restrepo (2020) and Dauth et al. (2021), Giuntella and Wang (2019) found significant negative effects of robot exposure on the employment and wage level of workers. With a similar index design, Du and Wei (2021) demonstrated that robot exposure results in a higher local unemployment rate in the short run, but this negative impact is reversed in the long run. As a global leading robot usage country, China demands more theoretical and empirical research in understanding how robots and human labor interact with each other; this research provides more evidence from the perspective of migrant workers.

Almost all previous studies focused on the whole local labor market, neglecting the fact that different social groups make up the labor force. Due to their special social status, some groups may be more vulnerable to the labor market shock caused by technological changes, such as women, ethnic minorities, and immigrants. Borjas and Freeman (2019) indicated that an additional robot is equivalent to two to three human workers, meaning that in the US, robots tend to better substitute for local labor than immigrants. As immigrant workers (usually low-educated) have been an indispensable part of the labor market in the US and European countries since the last century (Borjas, 1991; Borjas & Bratsberg, 1996; Gang & Rivera-Batiz, 1994), one may witness the significant influence of technological changes on these low-skilled immigrants in the near future, which is also the case for migrant workers in China. To the best of our knowledge, this research is the first to discuss the labor market shock of technological changes on special social groups, using migrant workers in China as a case study.

5.3.2. Robot exposure and migrant sorting

This research focuses on the impact of industrial robot exposure on migrants in China. As documented in Section 5.2, China has become the leading country in industrial robot usage, accounting for a large proportion of robot sales in the global market (Cheng et al., 2019). Thus, it is expected to witness a significant labor market shock induced by robot exposure, especially in highly automated industries. However, it is inappropriate to consider the effects of robot exposure on natives and migrants equivalently because of the *hukou* barriers, making migrant workers a distinct social group in China.

China's economic miracle is inseparable from the contribution of the large number of migrant workers. With the relaxation of migration restrictions since the 1980s, a large scale of rural idle labor marched into coastal cities rapidly. According to the *2018 China Migration Population Development Report*, there were nearly 250 million migrants in 2015, accounting for a large part of the labor market. However, because of the *hukou* barriers, migrants are not equivalent to urban citizens in terms of social welfare and services, such as education, insurance, and health care (Song, 2014; Wang & Hu, 2019). Moreover, migrants face severe *hukou* discrimination in the labor market, squeezing them into dirty, low-skilled jobs that native workers are unwilling to take (Meng & Manning, 2010; Song, 2016; Zhang, 2020). Therefore, migrants are more likely to perceive destination cities as places to make money instead of places to settle (Zhou & Tang, 2021). According to the rough calculation by Meng (2012), the average stay duration of migrant workers in the destination city is only seven years. Consequently, on the one hand, migrant workers are highly susceptible to the adoption of industrial robots in destination cities due to the characteristics of their jobs. On the other hand, they are more mobile to chase higher wages, making their reaction to the robot exposure faster and more intense.

As documented in Section 5.3.1, there are generally two opposite effects of robot exposure on labor demand. When the *displacement effect* dominates, industrial robots will replace migrant workers and squeeze them out of the occupation. This phenomenon is particularly significant in production occupations as they have more routine tasks and are easier to be replaced by robots. Therefore, exposure to robots will significantly decrease the proportion of migrant workers in production occupations. By contrast, when the *reinstatement effect* dominates, robot exposures will create new jobs to complement robots and increase the productivity level, thus expanding the economy and more labor demand. Therefore, exposure to robots will significantly increase the proportion of migrant workers in production jobs. As such, this research proposes first two antagonistic hypotheses to identify the effect that dominates the occupation selection of migrant workers:

H5.1a: Robot exposure decreases the proportion of migrant workers in production occupations when the *displacement effect* dominates.

H5.1b: Robot exposure increases the proportion of migrant workers in production

occupations when the *reinstatement effect* dominates.

Furthermore, according to Autor et al. (2003), occupation is highly associated with skill level. Within production occupations, industrial robots will replace low-skilled occupations with routine tasks and create new complementary occupations that usually require high education and skills. This adjustment of occupation structure will crowd out low- and mid-skilled migrant workers and attract more high-skilled migrant workers, leading to a positive skill sorting of migrants. Then, the replaced migrant workers will be reallocated to other manual-task occupations in other sectors with lower skill requirements, such as the service sector (Autor & Dorn, 2013; Autor & Salomons, 2018). In summary, robot exposure will lead to more low- and high-skilled migrant workers across all occupations, thereby making it similar to the skill polarization in developed countries. This research proposes second two parallel hypotheses to identify the skill selection induced by robot exposure:

H5.2a: Robot exposure increases the proportion of high-skilled migrant workers in production occupations, i.e., a positive skill sorting.

H5.2b: Robot exposure increases the proportion of both low- and high-skilled migrant workers across all occupations, i.e., a U-shaped skill sorting.

In summary, this research attempts to investigate the occupation and skill selection of migrants by robot exposure in urban China. The following sections propose empirical models to test these hypotheses.

5.4. Empirical method and data

5.4.1. Empirical design

This research uses two steps to discuss occupation and skill selection. First, we propose an individual-level binary probit model to explore how robot exposure influences the occupation selection of migrants, as shown below:

$$\Pr\left(Occupation_{p,c,t}\right) = \Phi\left(\alpha + \beta Robot\ Expospure_{c,t-1} + \boldsymbol{D\theta} + \boldsymbol{C\gamma} + \boldsymbol{\delta} + \boldsymbol{\sigma} + \varepsilon_{p,c,t}\right) \tag{5.1}$$

where $Occupation_{p,c,t}$ is a dummy variable indicating the occupation of migrant p in city c and year t; $Robot\ Expospure_{c,t-1}$ denotes the condition of robot exposure in city c and year t-1; $\boldsymbol{D}$ and $\boldsymbol{C}$ are vectors of controlling variables capturing migrant p's demographic and city-specific characteristics, respectively; $\boldsymbol{\delta}$ and $\boldsymbol{\sigma}$ are time and origin region fixed effects, respectively; Φ is the cumulative distribution function of the normal distribution. The key coefficient of this equation is β, which denotes the impact of robot exposure on migrants' occupation decisions. As there is a certain delay in the response of migrants to robots, this research introduces a one-year lagged robot index into the model. As $Occupation_{p,c,t}$ indicates production occupations, if $\beta < 0$ and is statistically significant, robot exposure leads to fewer migrants in production occupations, suggesting that the displacement effect dominates and H5.1a holds. By contrast, if $\beta > 0$ and is statistically significant, robot exposure leads to more migrants in production occupations, suggesting that the reinstatement effect dominates and H5.1b holds.

Second, an OLS estimation is employed to explore the effects of robot exposure on the skill level of migrants within and across occupations, as shown below:

$$Skill_{p,c,t} = \alpha + \beta Robot\ Expospure_{c,t-1} + \boldsymbol{D\theta} + \boldsymbol{C\gamma} + \boldsymbol{\delta} + \boldsymbol{\sigma} + \varepsilon_{p,c,t} \quad (5.2)$$

Where $Skill_{p,c,t}$ denotes the skill level of migrant p in city c and year t; the definitions of other variables are the same as those in Equation (5.1). Similarly, the key variable of this equation is β, which denotes the effects of robot exposure on the skill level of migrants. This research uses the education level to capture the skill level of migrants. If H5.2a holds, $\beta > 0$ is expected, suggesting a positive educational selection of migrants due to robot exposure.

However, this equation cannot capture the nonlinear skill sorting of migrants. Therefore, a multinomial logit model is employed to estimate the heterogeneous effects of robot exposure on different skill levels, as shown below:

$$\Pr\left(Skill\_dum_{p,c,t} = i\right) = \begin{cases} 1/\left(1 + \sum_{i=2}^{\boldsymbol{I}} \exp \boldsymbol{X_{p,c,t}\delta_i}\right) & i = 1 \\ \exp \boldsymbol{X_{p,c,t}\delta_i}/\left(1 + \sum_{i=2}^{\boldsymbol{I}} \exp \boldsymbol{X_{p,c,t}\delta_i}\right) & i > 1 \end{cases} \quad (5.3)$$

where $Skill\_dum_{p,c,t}$ denotes the skill level dummy of migrants (*i*=1, 2, 3, 4, and 5), including primary school and below, junior high school, senior high school, three-year college, as well as four-year college and above; $\boldsymbol{X_{p,c,t}}$ denotes the vector of variables, which are similar to those of the above equations; $\boldsymbol{\delta_i}$ denotes the coefficients vector of influencing variables in skill level i. In this equation, the key variable β_i (included in $\boldsymbol{\delta_i}$) ranges in different skill levels and thus can reflect the heterogeneous effects of robot exposure. If H 5.2b holds, a significantly larger β_i in low- and high-skill levels than mid-skill levels is expected.

5.4.2. Robot exposure measure

This research mainly focuses on industrial robots, referring to "actuated mechanism programmable in two or more axes with a degree of autonomy, moving within its environment, to perform intended tasks" (ISO 8373:2012). Notably, only "multipurpose industrial robots" are considered. From 1993 to date, the International Federation of Robotics (IFR) has collected detailed installation (in the current year) and stock number of robots by country, industry, and year, covering 50 countries. Unfortunately, this database does not cover robot adoption at the regional level of the countries. To demonstrate the local robot exposure at the prefectural-city level, this research follows the Bartik-style measure proposed by Acemoglu and Restrepo (2020) and Dauth et al. (2021).

The measure constructs the local robot exposure index using the weighted average of robot installation by employment share of each industry in a local labor market, as shown below:

$$Robot\ Exposure_{ct} = \sum_{d=1}^{D} \left(\frac{emp_{d,c,t=base}}{emp_{c,t=base}} \times \frac{Robot_{d,t}}{emp_{d,t=base}} \right) \tag{5.4}$$

where $emp_{d,c,t}$ represents the employment level in industry d, city c, and year t; $Robot_{d,t}$ represents the installation units of industrial robots in industry d and year t from the IFR database. The basic idea of this measure comprises the following two assumptions: (1) the robot penetration degree in each region within a country is equivalent, and (2) the employment distribution by industry does not change significantly starting from a specific base year. Accordingly, this measure predicts the robot adoption condition at the city level based on the nationwide robot adoption condition and employment share by city and industry.

The employment data is derived from the second *China Enterprise Economic Census* in 2008, which provides a detailed number of workers by industry and city. Because there are differences in the industrial classifications of China and IFR, this research groups the two standards into 19 sectors: agriculture, forestry, fishing; mining and quarrying; food and beverages; textiles, wood and furniture; paper and printing; plastic and chemical products; minerals; basic metals; metal products; electrical and electronics; industrial machinery; automotive; other vehicles; other manufacturing production; electricity, gas, water supply; construction; education, research, and development; and other services. Following Equation (5.4), this research aggregates employment and robot installation/stock based on this industry classification and calculate the robot exposure index across Chinese prefecture-level cities. Figure 5.3 depicts the difference in the city-level robot exposure (stock) in 2008 and 2019, demonstrating the robot penetration conditions during this period. The figure reveals that the eastern coastal cities significantly adopt more robots than inland cities, except the *Hubei* province. The agglomeration of manufacturing industries in coastal areas and *Hubei* province may be responsible for this uneven distribution (Cheng et al., 2019).

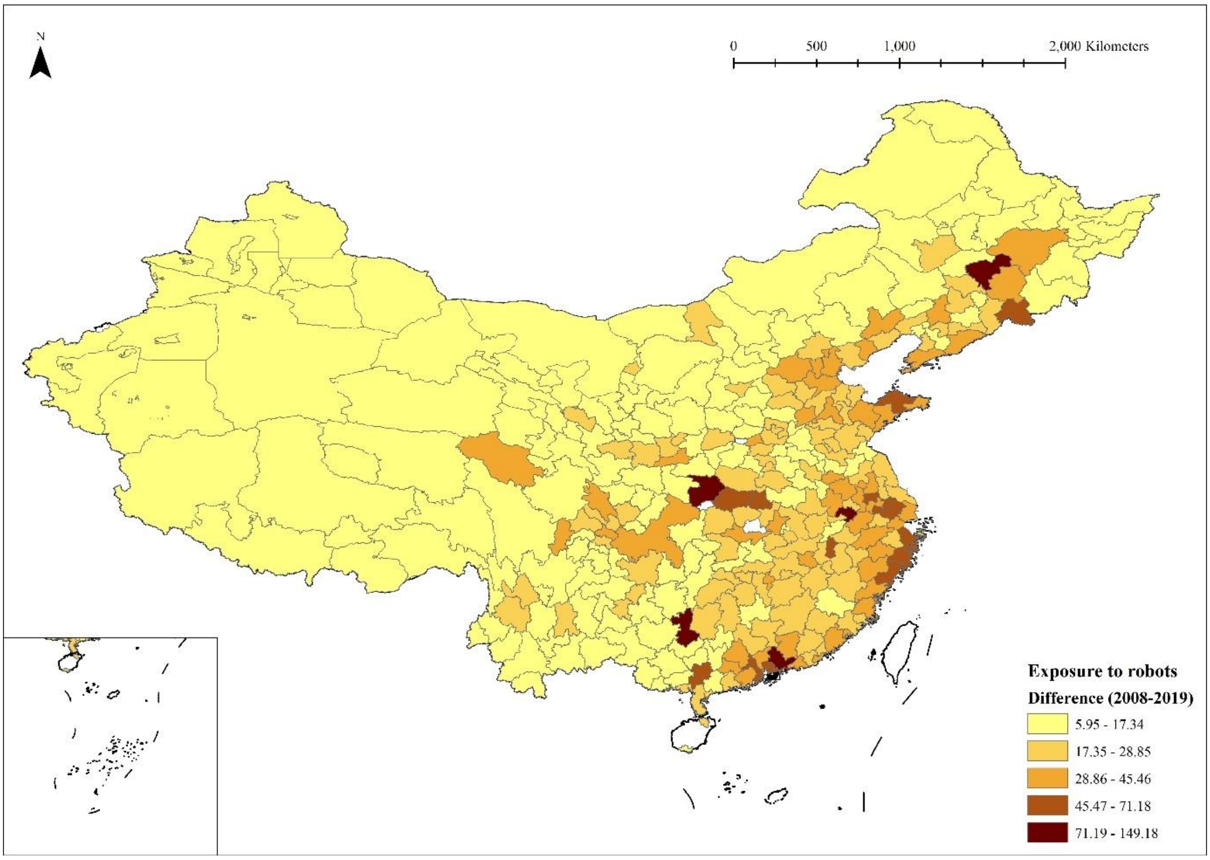

Figure 5.3 Robot exposure (stock difference) across Chinese prefecture-level cities, 2008-2019

However, there is an endogeneity concern that the robot exposure index is correlated with some omitted variables. For example, the original regional agglomeration of manufacturing industries influences the skill sorting of migrants and may lead to more robot adoption simultaneously. To address this concern, this research first introduces more controlling variables to capture the original industry development condition, which will be discussed in the following subsection. Additionally, this research proposes an IV to eliminate the endogeneity concern. As the adoption of industrial robots relies heavily on imports from other developed countries, the robot exposure condition in these countries influences that in China, but it does not exert influence on the labor market of migrants in China. Therefore, this research recalculates the local robot exposure index using the robot installation data about six developed countries—France, Sweden, Denmark, Finland, Italy, and the US. Then, this research takes the average of the calculated six robot exposure indices and use it as the IV.

5.4.3. Other data source

The migrant data is derived from the *China Dynamic Migrant Survey* (CMDS) 2011–2017. Since 2009, the China national health commission has conducted an annual, cross-sectional, and national representative survey targeting the "floating population" (migrants without local *hukou*), covering nearly 200,000 migrants in 31 provinces or municipalities each year. This survey provides detailed information about household heads' demographics, migration, and employment. This research focuses on inter-city migrant workers still in the urban labor market. Therefore, the research sample select migrants aged 15–65 in urban areas that have moved from another prefectural city.

The CMDS only includes migrants without local *hukou*, neglecting permanent migrants who have already obtained local *hukou*. Due to the skill-biased selection feature of the *hukou* system, this sample selection decreases the proportion of high-skilled migrants; thus, the skill distribution of migrants will be biased. To address this issue, the research sample only considers new migrants who arrive at the destination city one year before the survey year. The idea is that migrants usually take a certain amount of time to meet the requirements of the *hukou* obtainment application, which usually exceeds a year. The research also compares the results of new and all migrants. The statistical description of new migrants is presented in Table 5.1.

Table 5.1 The statistical description of migrants

Variable	Observations	Mean	Std. Dev.	Min	Max
Years of schooling	180,618	10.423	3.011	0	19
Occupation (production=1)	180,618	0.197	0.398	0	1
Age	180,618	31.051	9.514	15	65
Gender (male=1)	180,618	0.532	0.499	0	1
Marital status (married=1)	180,618	0.664	0.472	0	1
Hukou status (urban=1)	180,618	0.182	0.386	0	1
Same province (same=1)	180,618	0.429	0.495	0	1

Moreover, this survey provides the education level of migrants. The education level is taken as their skill levels after being converted into years of schooling. The CMDS also records migrants' occupation information with a rough classification. Thus, it is unable to do a task-based occupation analysis. Instead, this research classifies all occupations into two—production (including manufacturing, transportation, construction, and other assisting occupations) and service (including commerce, catering, manual service, and public service occupations). In general, migrants in service occupations account for approximately 80% of all migrants, as presented in Table 5.1.

In addition to the demographic controls, this research introduces city-level controlling variables obtained from the China City Yearbook, 2011–2017. First, because large cities with more population and higher economic development levels tend to attract more migrants (especially high-skilled), this research includes *Urban population* and *GDP per capita* to capture these city features. Second, as migrants also incur living costs in the destination cities, this research includes *housing price*, *Student/teacher ratio*, *College student per capita*, and *Doctor per capita* to capture migration costs. Finally, as discussed in Section 5.4.2, the industry development condition simultaneously influences robot adoption and migrant skill distribution. To address this concern, this research includes a series of variables to control the industry development condition, including *FAI ratio*, *Industry structure, Manufacturing employment ratio,* and *Firm density.* The definition and calculation of all variables are presented in Table 5.2.

Table 5.2 Description of city-level controlling variables

Variable	Description
Urban population	The number of populations living in urban areas.
GDP per capita	The GDP per capita of urban residents (Yuan)
Housing price	The average transaction price of commercial housing (Yuan)
Student/teacher ratio	The ratio of the number of students to teachers in primary and junior high schools (compulsory education).
College student per capita	The number of college students per 10,000 people.
Doctor per capita	The number of qualified doctors per 10,000 people.
FAI ratio	The ratio of foreign direct investment amount to GDP.
Industry structure	The ratio of the GDP proportion of the tertiary industry to the proportion of GDP of the secondary industry.
Manufacturing employment share	The employment ratio of manufacturing industries.
Firm density	The number of industrial enterprises above the designated size per 10,000 workers.

5.5. Empirical results

5.5.1. Occupation sorting of migrants

This subsection first discusses the occupation sorting of migrants; probit model results are presented in Table 5.3. Column (1) includes the key variable *Robot exposure* and time and location dummy variables, which demonstrate that robot exposure has significant positive effects on the probability of migrants in production occupations. After introducing demographic and city-level controlling variables in Columns (2) and (3), respectively, the significance and sign of robot exposure are still stable with a slight decrease in the value. Column (4) only focuses on production occupations in manufacturing industries with the same controls employed in Column (3), which has a similar result that robot exposure significantly affects the probability of migrants in manufacturing occupations. Columns (5) and (6) of Table 5.3 compare all migrants with new migrants in previous columns. The results are still highly stable with slightly larger coefficients of the key variable, indicating that cities with more robot adoption are more attractive to long-term migrants.

Table 5.3 Occupation sorting of migrants: probit model estimation

	New migrants				All migrants	
	(1)	(2)	(3)	(4)	(5)	(6)
	Production	Production	Production	Manufacturing	Production	Manufacturing
Robot exposure	0.022***	0.027***	0.011**	0.013**	0.017***	0.022***
	(0.003)	(0.004)	(0.005)	(0.005)	(0.002)	(0.003)
Age		0.024***	0.024***	0.000	0.036***	0.014***
		(0.003)	(0.003)	(0.004)	(0.002)	(0.002)
Square of age		-0.000***	-0.000***	-0.000	-0.000***	-0.000***
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Gender (Male=1)		0.517***	0.507***	0.076***	0.508***	0.088***
		(0.007)	(0.008)	(0.010)	(0.004)	(0.005)
Marital status (married=1)		-0.071***	-0.081***	-0.136***	-0.086***	-0.152***
		(0.010)	(0.011)	(0.013)	(0.007)	(0.008)
Urban hukou		-0.328***	-0.331***	-0.324***	-0.228***	-0.256***
		(0.011)	(0.011)	(0.015)	(0.006)	(0.008)
Same province		-0.056***	-0.035***	-0.037**	0.038***	0.026***
		(0.010)	(0.011)	(0.016)	(0.006)	(0.008)
ln (Urban population)			-0.093***	-0.024*	-0.069***	-0.010
			(0.010)	(0.012)	(0.005)	(0.007)
ln (GDP per capita)			-0.040***	-0.038**	0.021***	0.024**
			(0.015)	(0.019)	(0.008)	(0.010)
ln (Housing price)			-0.052***	-0.131***	-0.070***	-0.125***

			(0.019)	(0.024)	(0.010)	(0.013)
Student/teacher ratio			0.005***	0.007***	0.003***	0.005***
			(0.002)	(0.002)	(0.001)	(0.001)
ln (College students per capita)			-0.017**	-0.022**	-0.025***	-0.060***
			(0.007)	(0.009)	(0.004)	(0.005)
ln (Doctor per capita)			0.134***	0.196***	0.073***	0.104***
			(0.018)	(0.023)	(0.009)	(0.013)
FAI ratio			-0.075***	-0.202***	-0.069***	-0.172***
			(0.023)	(0.037)	(0.011)	(0.018)
Industry structure			-0.182***	-0.322***	-0.139***	-0.212***
			(0.014)	(0.020)	(0.007)	(0.010)
Manufacturing employment share			0.366***	0.397***	0.190***	0.211***
			(0.057)	(0.054)	(0.014)	(0.012)
Firm density			0.002***	0.002***	0.002***	0.002***
			(0.000)	(0.000)	(0.000)	(0.000)
Constant	-2.131***	-2.712***	-0.997***	-0.979**	-1.575***	-1.372***
	(0.163)	(0.171)	(0.250)	(0.420)	(0.136)	(0.213)
Survey year FE	Yes	Yes	Yes	Yes	Yes	Yes
Original region FE	Yes	Yes	Yes	Yes	Yes	Yes
Destination region FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	180,609	180,609	161,169	161,136	596,511	596,497
Pseudo R^2	0.090	0.126	0.138	0.193	0.120	0.166

Notes: 1. Robust standard errors are shown in parentheses; 2. ***, **, and * denote significance at the 1, 5, and 10 percent levels, respectively.

Statistically, the research calculates the marginal effects of robot exposure in Column (3), showing that adopting one robot per 10,000 workers improves 0.28% of the probability of migrants working in production occupations. All the results indicate that robot exposure increases the migrant workforce in production occupations instead of service occupations, implying that robot adoption has created more jobs in production occupations, compensating the job it replaces for migrants. Put differently, the reinstatement effect dominates in the process of robot exposure influencing the labor market of migrants, supporting **H5.1b** instead of **H5.1a**.

Apart from the key variable, other variables also have significant effects on the occupation distribution of migrants. Regarding the demographic variables, results demonstrate that young, male, unmarried, and interprovincial migrants with rural *hukou* are more likely to undertake production occupations. Regarding city-level variables, big cities, characterized by more urban population, higher GDP per capita, higher housing price, more college students, and less education and medical resource per capita, are more likely to attract migrants in service sectors, which is consistent with the reality. Regarding the industrial conditions variables, cities with less FAI, less GDP ratio of tertiary industry, more manufacturing employment share, and more industrial firms are more likely to attract migrants in production occupations.

The results presented in Table 5.3 may have the endogenous problem as cities with more manufacturing firms have higher possibilities to adopt industrial robots. This problem may still exist even if the empirical model introduces several variables about industrial development conditions. Therefore, this research introduces an IV to solve this problem, as presented in Table 5.4. Columns (1) and (3) of Table 5.4 show the first stage results of the IV estimation. The IV has significant effects on robot exposure, confirming its reliability. Columns (2) and (4) of Table 5.4 show the second stage results for new migrants, revealing that the coefficient of robot exposure is nearly equal to that in the probit regression for migrants in all production occupations, whereas the coefficient of robot exposure is slightly larger than that in the probit regression for migrants undertaking production occupations in manufacturing industries. Similarly, Columns (5) and (6) of Table 5.4 indicate that the second stage IV estimation for all migrants is larger, but the coefficient of the key variable is still significant. As such, IV estimations confirm the reliability of the results.

Table 5.4 Occupation sorting of migrants: probit model + IV estimation

	New migrants				All migrants	
	Production		Manufacturing		Production	Manufacturing
	(1)	(2)	(3)	(4)	(5)	(6)
	First-stage	Second-stage	First-stage	Second-stage	Second-stage	Second-stage
IV_Robot exposure	0.419***		0.419***			
	(0.003)		(0.003)			
Robot exposure		0.012***		0.041***	0.025***	0.042***
		(0.006)		(0.007)	(0.003)	(0.003)
Demographic controls	Yes	Yes	Yes	Yes	Yes	Yes
City-level controls	Yes	Yes	Yes	Yes	Yes	Yes
Survey year FE	Yes	Yes	Yes	Yes	Yes	Yes
Original region FE	Yes	Yes	Yes	Yes	Yes	Yes
Destination region FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	161,169	161,169	161,136	161,136	596,511	596,497

Notes: 1. Robust standard errors are shown in parentheses; 2. ***, **, and * denote significance at the 1, 5, and 10 percent levels, respectively.

Robust results confirm that more migrants flow into production occupations instead of service occupations, which is inconsistent with the evidence found in developed countries (Autor & Dorn, 2013; Autor & Salomons, 2018). However, this result cannot exclude one possible reason that the replaced migrant workers may not flow into the local service sector but will move to other cities. These substituted migrants can find other production occupations with similar job requirements in cities with less robot exposure. It is also possible that they will move into the service sector in big cities with higher demand for service and higher wage levels (Eeckhout et al., 2014). As discussed above, results confirm that big cities disproportionately attract migrants into service industries, which partially indicates that the displaced workers may have moved into the service sectors in other big cities instead of staying in the local labor market.

5.5.2. Skill sorting of migrants

This subsection then discusses the skill sorting of migrants, employing two empirical models. Table 5.5 presents the OLS estimation of how robot exposure influences the schooling years of migrants. Column (1) only includes the key variable as well as time and location dummy variables. The result shows that robot exposure significantly improves the educational level of migrants. Then, demographic and city-level controlling variables are introduced shown in Columns (2) and (3), respectively. The key variable, robot exposure, is still significant and positive. Statistically, Column (3) suggests that adopting one robot per 10,000 workers increase the schooling years of migrants by 0.024. In Column (4), the model only considers migrants in production occupations, including manufacturing, transportation, and construction, but the key variable is not significant. In Column (5), the model further limits the sample to migrants with production occupations in manufacturing industries. results demonstrate the considerable effects of positive skill sorting on migrants by robot exposure, supporting H5.2a. This result also reveals that robot exposure mainly induces skill sorting in manufacturing industries, whereas the high degree of routine-type tasks in these industries is responsible for this consequence. The result of all migrants reveals similar significant positive effects of robot exposure, as shown in Column (6).

Table 5.5 Skill sorting of migrants: OLS estimation

	New migrants					All migrants
	All			Production	Manufacturing	All
	(1)	(2)	(3)	(4)	(5)	(6)

	Years of schooling	Years of schooling	Years of schooling	Years of schooling	Years of schooling	Years of schooling
Robot exposure	0.075*** (0.007)	0.049*** (0.006)	0.024*** (0.007)	0.005 (0.015)	0.045** (0.022)	0.027*** (0.003)
Demographic controls	No	Yes	Yes	Yes	Yes	Yes
City-level controls	No	No	Yes	Yes	Yes	Yes
Survey year FE	Yes	Yes	Yes	Yes	Yes	Yes
Original region FE	Yes	Yes	Yes	Yes	Yes	Yes
Destination region FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	180,617	180,617	161,177	32998	16352	596,512
Adjusted R^2	0.088	0.298	0.291	0.258	0.241	0.319

Notes: 1. Robust standard errors are shown in parentheses; 2. ***, **, and * denote significance at the 1, 5, and 10 percent levels, respectively.

The endogenous problem also exists in the skill sorting model, as cities with more manufacturing industries tend to have more low-skilled labor and are more likely to adopt industrial robots. To solve the endogenous problem, the research introduces the same IV into the model, and the results are presented in Table 5.6. Column (1) presents the first stage results of the IV estimation, and IV still has significant effects on the key variable, robot exposure. Column (2) presents the second stage results of the IV estimation for new migrants, demonstrating that the coefficient of robot exposure is smaller than that in the OLS estimation but is still significantly positive at least at the 10% level. The empirical model further limits the sample to migrants with production occupations and migrants with production occupations only in manufacturing industries in Columns (3) and (4), respectively. The results are highly stable and similar to the OLS estimation, revealing that robot exposure only induces skill sorting of migrants with production occupations only in manufacturing industries. Finally, Column (5) presents the results of all migrants, which are also similar to the OLS estimation. These results confirm the stability of the OLS estimation.

Table 5.6 Skill sorting of migrants:2SLS + IV estimation

	New migrants				All migrants
	All occupations		Production	Manufacturing	All occupations
	(1)	(2)	(3)	(4)	(5)
	First-stage	Second-stage	Second-stage	Second-stage	Second-stage
IV_Robot exposure	0.419***				
	(0.003)				
Robot exposure		0.017*	0.011	0.077**	0.017***
		(0.009)	(0.022)	(0.039)	(0.004)
Demographic controls	Yes	Yes	Yes	Yes	Yes
City-level controls	Yes	Yes	Yes	Yes	Yes
Survey year FE	Yes	Yes	Yes	Yes	Yes
Original region FE	Yes	Yes	Yes	Yes	Yes
Destination region FE	Yes	Yes	Yes	Yes	Yes
Observations	161,177	161,177	32,998	16,352	596,512
Adjusted R^2	0.882	0.291	0.258	0.241	0.319

Notes: 1. Robust standard errors are shown in parentheses; 2. ***, **, and * denote significance at the 1, 5, and 10 percent levels, respectively.

However, the linear regression cannot reveal the nonlinear skill sorting of migrants, which requires using the multinomial logit regression to compare the heterogeneous effects on migrants with different education levels, as presented in Table 5.7. Panel A includes migrants in all types of occupations. Using migrants with a primary school education or below as a reference, results reveal that as robot exposure increases, the migrants' education level also increases. However, this increase is not linear, but it has a vertex at three-year college. This tendency is more significant within all production occupations and production occupations in manufacturing industries, as presented in Panels B and C, respectively. These results imply that robot exposure induces positive skill sorting across all occupations, especially in production occupations. As three-year colleges are usually vocational schools that aim to cultivate the manufacturing workforce, migrants with three-year college degrees are disproportionately attracted to cities with more robot adoption. Lastly, Panel D considers migrants in service occupations, which still reveals a (relatively less) significant positive skill sorting by robot exposure. Thus, this research does not observe replaced low-skilled labor flow into local service sectors.

Table 5.7 Nonlinear skill sorting of migrants by occupation: multinominal logit regression

	Education categories (reference: primary school and below)			
	Junior high school	Senior high school	Three-year college	Four-year college and above
	Panel A: All occupation			
Robot exposure	0.083***	0.112***	0.124***	0.114***
	(0.010)	(0.011)	(0.013)	(0.015)
Observations	161,178			
Pseudo R^2	0.131			
	Panel B: production occupation (including manufacturing, transportation, and construction)			
Robot exposure	0.075***	0.127***	0.186***	0.131***
	(0.020)	(0.022)	(0.029)	(0.043)
Observations	32,998			
Pseudo R^2	0.130			
	Panel C: production occupation (only including manufacturing)			
Robot exposure	0.054*	0.140***	0.268***	0.027
	(0.029)	(0.033)	(0.042)	(0.079)
Observations	16,352			
Pseudo R^2	0.127			
	Panel D: service occupation (excluding occupations in public and research sectors)			
Robot exposure	0.076***	0.100***	0.082***	0.143***
	(0.015)	(0.016)	(0.019)	(0.023)
Observations	87,042			
Pseudo R^2	0.113			
Demographic controls	Yes	Yes	Yes	Yes
City-level controls	Yes	Yes	Yes	Yes
Survey year FE	Yes	Yes	Yes	Yes

Notes: 1. Robust standard errors are shown in parentheses; 2. ***, **, and * denote significance at the 1, 5, and 10 percent levels, respectively.

In summary, empirical results support H5.2a, indicating that migrants are positively sorted within production occupations, but are inconsistent with H5.2b, showing that robot exposure increases the proportions of both low- and high-skilled migrant workers across all occupations. However, similar to H5.1, the rejection of H5.2b cannot rule out the possible reason that the displaced low-skilled migrants may have flowed into the service sectors in other cities, especially the big cities demanding more services. Eeckhout et al. (2014) proposed that in addition to high-skilled labor, big cities also attract extremely low-skilled workers into service sectors to complement the high-skilled ones. This is also the case in China, where the manufacturing industry workers mainly concentrate on mid-scale cities, whereas big cities have the largest proportion of service industry labor (see Figure 5.4). As cities with high degrees of robot exposure tend to be the mid-scale ones, displaced migrants cannot re-enter the small local service sector but flow to other big cities. Therefore, H5.2b may not hold at the city level but hold at the national level. Data about inter-city migration flows by skill are required to test this hypothesis, which is still relatively difficult to find.

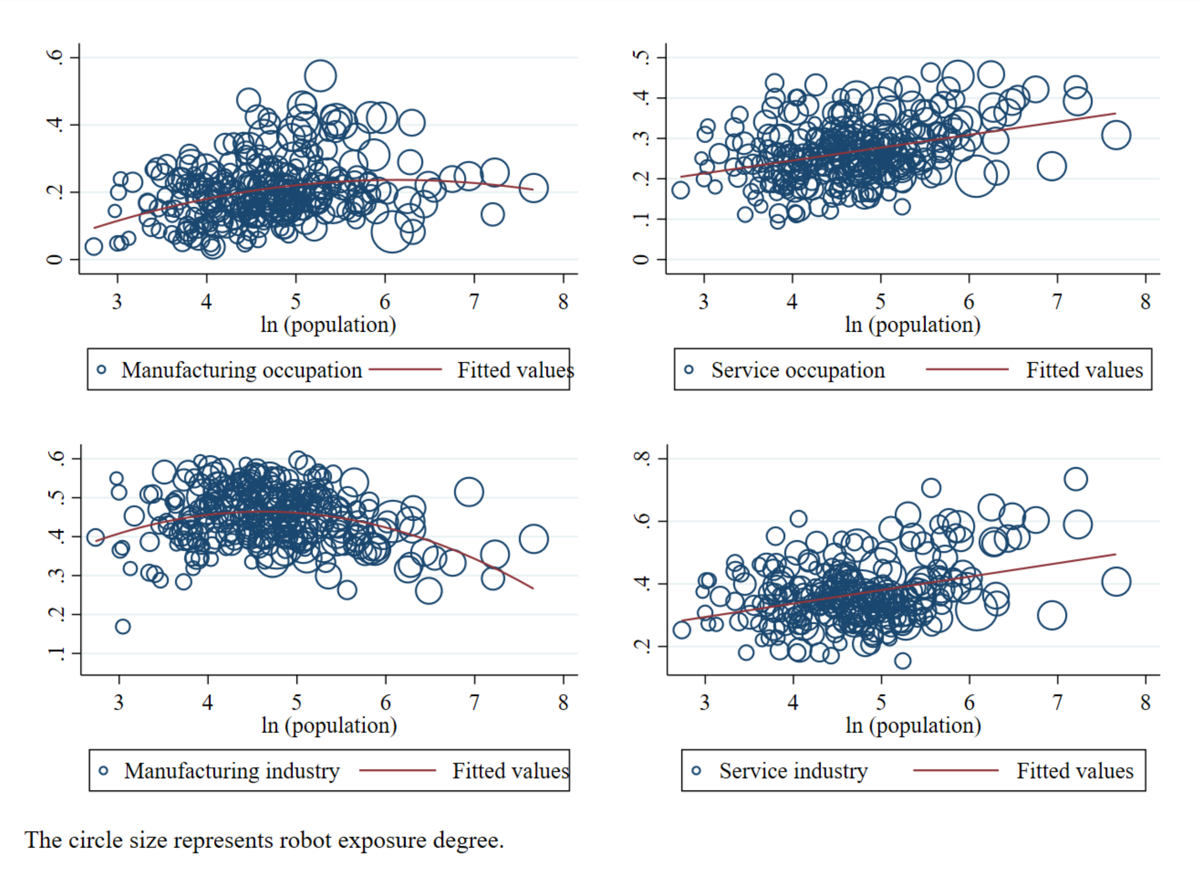

Figure 5.4 Employment share by occupation/industry

Note: the data is calculated by the author according to 2015 China 1% population sampling survey data.

5.6. Conclusions and discussions

This chapter mainly discusses how robot exposure influences the occupation and skill sorting of migrants in urban China, contributing to the heated debate on how technological change reshapes the labor market, which achieves Objective 2.3. This chapter proposes that industrial robot adoption will attract more high-skilled migrants in production sectors but replace and crowd out low-skilled migrants into service sectors. This chapter first constructs a predicted robot exposure index to capture the city-level robot adoption condition based on the nationwide robot adoption and employment share by industry and city. Then, based on the CMDS data from 2011 to 2017, This chapter develops empirical strategies to discuss migrants' occupation and skill sorting.

Empirical results reveal that robot exposure booms the local manufacturing industries, thus increasing the proportion of migrants working in production occupations, supporting H5.1b that the *reinstatement effect* dominates the robot exposure impact. Furthermore, these attracted migrants are mainly high-skilled labors, especially those with three-year college degrees, implying that industrial robots have replaced low-skilled migrants but attracted more high-skilled migrants to complement their tasks, supporting H5.2a. However, this research does not find evidence that the replaced migrants have flowed into the local service sector, failing to support H5.2b. These results provide a new perspective (i.e., migrant workers) to see how technological change reshapes the labor market, especially in emerging economies.

Moreover, empirical results are consistent with the evidence found in other countries that industrial robots substitute low-skilled labor (Acemoglu & Restrepo, 2020; Dauth et al., 2021) but inconsistent with the theory proposed by Autor et al. (2003) that automation will replace mid-skilled labor and disproportionately lead to more labor at the two extremes of the skill distribution. Three reasons may explain this difference. First, industrial robot is one of the technological changes and may only target low-skilled routine jobs, thus leading to significant positive sorting of migrants. Second, industrial robot is at the initial stage of development and can only have comparative advantages over low-skilled labor. Based on the China Employer-Employee Survey (CEES), Cheng et al. (2019) proposed that China is still at the industrial stage of replacing manual, dirty, and health-hazardous tasks by adopting robots. Finally, replaced migrants may have flowed into service sectors in other big cities demanding more services and providing higher wage levels than the current city. Therefore, Autor's theory may hold at the national level but is not necessarily true at the regional level.

These results raise serious concerns for migrant workers facing technological change, providing policy implications for the government to promote industrial upgrade and automation. As migrant workers have already endured plenty of surviving pressure in their destinations (Chan, 2012; Meng, 2012), industrial robots have undoubtedly caused enormous challenges for their limited work chances, squeezing and crowding them out into service sectors (locally or not locally) with lower salaries and requiring even lower skill levels. Policymakers should take proactive steps to solve this problem, which is underway as revealed by results, especially when the current localized *hukou* system is unable to provide sufficient social security. After decades of development, there has been a substantial transformation in the way works get done across industries rather than all jobs being entirely replaced by computers. Thus, as industrial robots and even artificial intelligence technology are beginning to pervade global economies, companies and the government should train the existing workforces in complementary works for machines, such as machine operation and maintenance training. Workers facing technological change should be granted a path to a new-style job at the current company or a skill that can direct them to a new company. Automation has provided a productivity-enhancing opportunity for China at a time when its demographic dividend is declining and a chance to avoid its drawbacks.

CHAPTER 6. UNAFFORDABLE HOUSING COSTS AND HETEROGENEOUS CROWDING-OUT EFFECTS ON MIGRANTS

6.1. Introduction

Differing from Chapters 4 and 5, this chapter will discuss the impact of migration costs on migrant selection. Significant migration costs occur in international migration, such as policy restrictions, transportation costs, and assimilation costs due to cultural differences. In contrast, migrants face different costs in internal migration. Chapter 3 has discussed how the *hukou* system raises significant living costs for migrants. This chapter will discuss another expensive migration cost, housing costs, and thus fulfil Objective 2.4.

This chapter first introduces housing costs into the self-selection framework of migration and conducts a theoretical analysis. The theory proposes that unaffordable housing prices at the destination discourage low-skilled migrants more significantly than high-skilled migrants (i.e., positive selection). Then, this chapter employs China Migrants Dynamic Survey (CMDS) database and other data sources to develop an empirical model to test proposed hypotheses, using the education level to represent personal skill level.

The rest of the chapter is as follows. Section 6.2 gives a research background of housing prices and migration in China. Section 6.3 reviews the literature about housing prices and migration.

Section 6.4 conducts a theoretical analysis. The effects of housing prices on the migrant selection are estimated empirically in Sections 6.5 and 6.6. Finally, the discussion and conclusion are presented in Sections 6.7 and 6.8, respectively.

6.2. Research background

Since the housing marketization and commercialization reform in 1998, housing prices have entered a fast lane in many cities in China, significantly contributing to the economic growth of the regions. However, as shown in Figure 6.1, this wave of housing price appreciation varies greatly depending on different tiers of cities. From 2002 to 2017, housing prices in the four first-tier cities, Beijing, Shanghai, Guangzhou, and Shenzhen, increased by 6.4 times, which is 4.9 times that of the national average in 2017. The housing price gap between quasi-first- and second-tier cities and other cities is also gradually widening. Undoubtedly, such high prices have caused severe housing affordability problems in recent years, making residents more likely to suffer from longer commuting time, less housing space, and worse housing conditions, especially in large cities (Chen et al., 2010; Li et al., 2020).

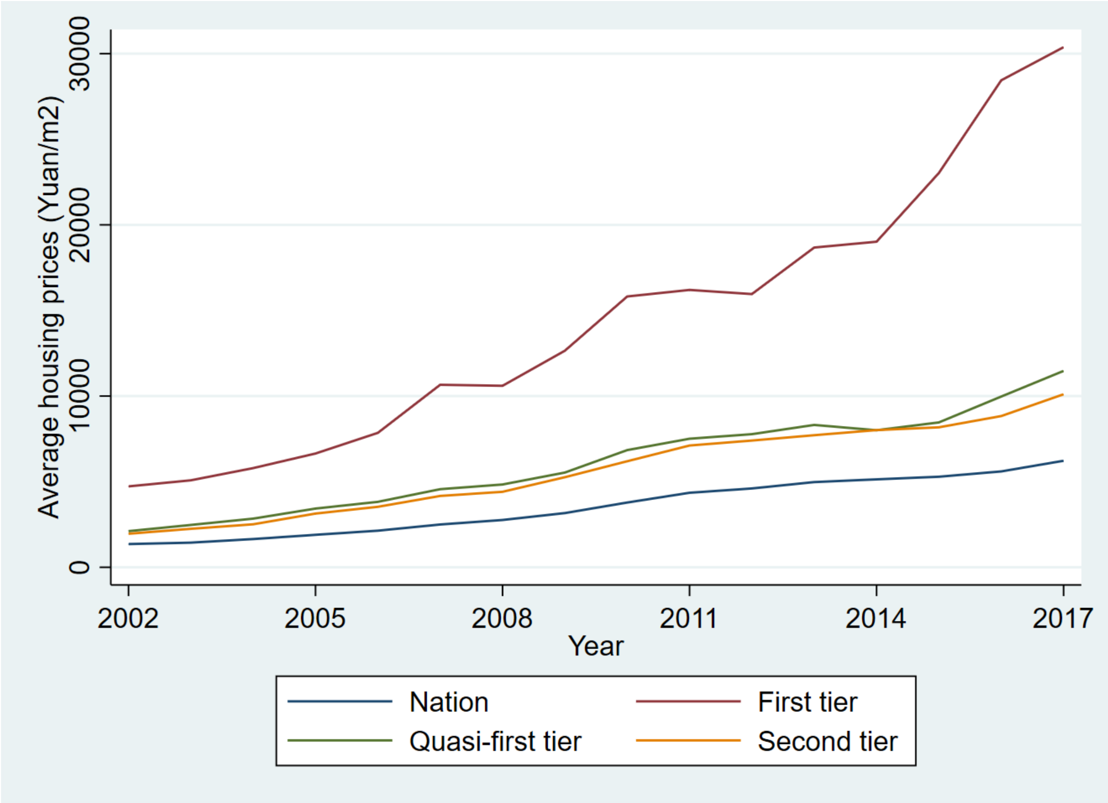

Figure 6.1 Average housing prices in different tiers of cities from 2002 to 2017.

Notes: 1. city ties are classified according to “China City Ranking” developed by New First-Tier City Research Institute, China Business Network (CBN); 2. the data comes from CEIC database.

Migrants are also severely hit by unaffordable housing prices since housing cost burdens have gradually become their most significant obstacle to survival and settlement in cities (Liu et al., 2017; Xie & Chen, 2018; Zang et al., 2015). Without the support of the public housing system, which *hukou* has placed restrictions on, migrants may suffer more severe housing issues than natives. Under these circumstances, some potential migrants may withdraw their intention to migrate, whereas migrants in destination cities may return home or move elsewhere. This process raises some crucial questions: “Who has been crowded out by unaffordable housing prices in destination cities?” or “Who keeps migrating into these cities?” These questions are essential to urbanization and economic development but are unanswered.

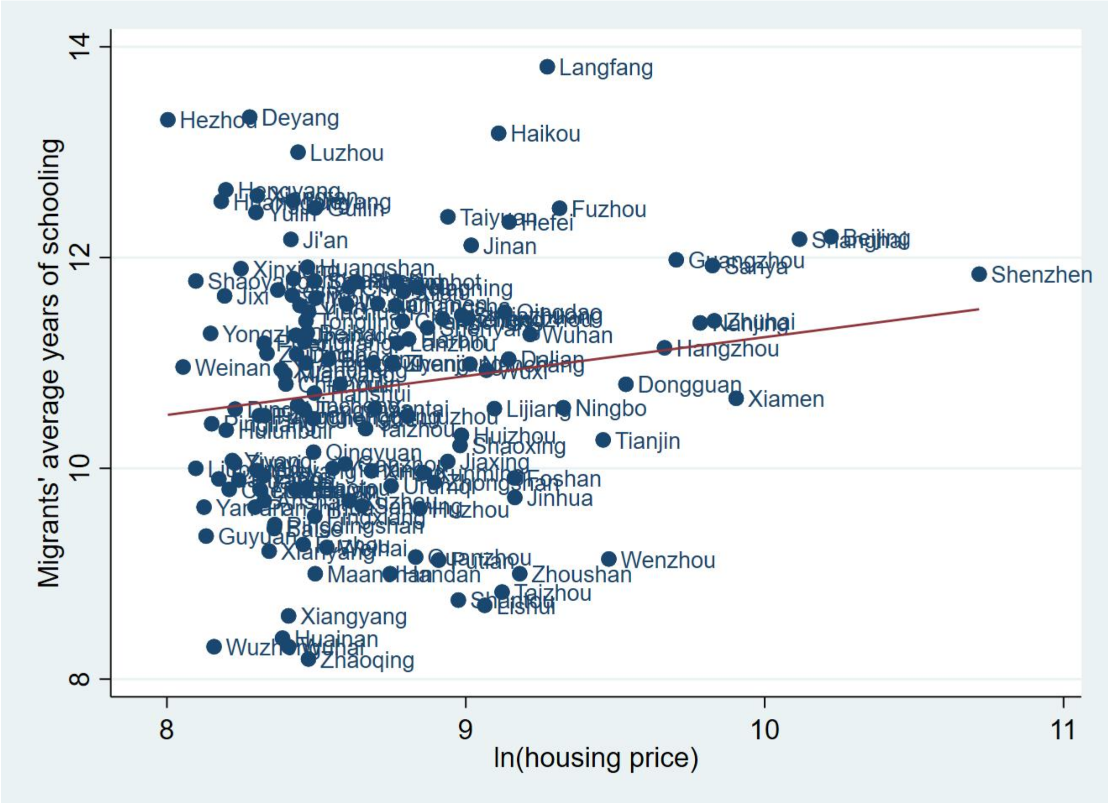

Figure 6.2 Relationship between housing prices and migrants' average years of schooling across cities in 2017

Notes: 1. cities with less than ten observations have been excluded; 2. the data comes from CMDS 2017 and CEIC database.

In China, Chen et al. (2019) attempt to answer how housing unaffordability selects migrants, and their results showed that superstar cities with extremely high housing prices are still the preferred destination of "elites" (with a postgraduate degree or above). However, they did not discuss other low-skilled migrants. To demonstrate the relationship between migrants' skill levels and housing prices, Figure 6.2 draws a scatter plot demonstrating that average migrants' schooling years increase significantly with urban housing prices, implying that higher housing prices select migrants with higher education levels into destination cities. However, the mechanism of this relationship and the selection degree of housing prices on migrant selection in China are still unclear, requiring more theoretical and empirical research. This chapter aims to investigate this mechanism and sheds light on the self-selection of migrants in China.

6.3. Literature review

6.3.1. Housing price and population migration

In the classical inter-urban equilibrium framework, income, housing costs, and amenity are the three most important factors influencing worker allocation (Roback, 1982; Rosen, 1979). Keeping amenity exogenous, the inflow of workers attracted by higher income levels will increase housing demand and drive up housing prices (Hui & Gu, 2009; Wang et al., 2017). This high housing price will, in turn, increase the cost of living, hinder the further entry of migrants, and finally reach equilibrium. Numerous empirical studies have identified adverse effects of housing prices on the inflow of migrants, compared with the incentives of income differentials (Berger & Blomquist, 1992; Gabriel et al., 1992; Michaelides, 2011; Rabe & Taylor, 2012). As this is an endogenous process, some scholars employed the simultaneous equation model to investigate the interaction between housing prices and population migration (Jeanty et al., 2010; Potepan, 1994). Their results indicate that higher net migration booms the housing price and, in turn, discourages further in-migration, highly consistent with the classical theory.

The role of housing prices in migration also varies for different cohorts. For homeowners, rising housing prices increase their wealth and promote their aspiration to live in other areas (Peng & Tsai, 2019; Zang et al., 2015). In contrast, falling housing prices will induce an equity lock-in effect of lower or even negative housing equity, essentially impeding their movements (Bloze & Skak, 2016; Bricker & Bucks, 2016; Chan, 2001; Foote, 2016). Moreover, Peng and Tsai (2019) argued that housing prices' impact on migration might be asymmetric. Based on the panel cointegration method, they found that housing prices positively influence migration in the long run, but the effect is not significant in the short run. For renters, they are not tied to housing and tend to have higher mobility (Berger & Blomquist, 1992; Kan, 2003). In such a case, rising housing prices will raise their housing expenditures and undermine first-time buyers' housing affordability.

Apart from the direct effect of housing price, some scholars have also paid attention to other housing attributes, including housing tenures and housing conditions. Lux and Sunega (2012) explored the effects of housing tenures on the intention of labor migration. They found that homeownership hampers the intention of migration in the case of becoming unemployed. Several Chinese scholars found that the housing condition of migrants in destination cities influences their long-term settlement decision and *hukou* transfer intentions, whereas poor housing conditions will promote the probability of return migration (Liu et al., 2017; Tao et al., 2015; Xie & Chen, 2018; Yang & Guo, 2018).

6.3.2. Housing price and migrant selection

The classical migration theory has successfully explained how income levels, housing prices, and amenities allocate workers but fails to give insight into the human capital divergence across cities. The past decades have witnessed the continuous gathering of highly skilled workers in big cities (Bacolod et al., 2021; Combes et al., 2008; De la Roca, 2017), partly explaining the higher wage levels in these cities. Previous scholars have proposed several explanations for this phenomenon, such as differences in industry composition across cities (Elvery, 2010; Hendricks, 2011), agglomeration effect (Berry & Glaeser, 2005), and other amenity differences (Diamond, 2016; Gyourko et al., 2013).

As households' primary expenditure, the housing cost also plays a significant role in differences in the skill composition of workers across cities. Gyourko et al. (2013) proposed that, in the United States, the rapidly rising housing price in "superstar cities" due to the increase in the number of high-income households and land use regulation has crowded out lower-income households and changed the local income distribution. Ganong and Shoag (2017) drew a similar conclusion that high housing prices in prosperous US cities have eroded the migration return of low-skilled workers and thus crowded them out. Broxterman and Yezer (2015) empirically identified a positive relationship between urban housing cost index and skill intensity ratio (the ratio of college-educated to those lacking a college degree), which means that more educated workers live in cities with higher housing costs. Based on micro-level and macro-level data in China, Chen et al. (2019) found that although elites still prefer "superstar cities," housing unaffordability has challenged the attractiveness of these cities.

In summary, although housing prices and skill composition of urban workers have attracted much attention from scholars, few studies have proposed direct evidence of how the housing cost selects migrants by skill. This chapter attempts to fill this research gap and investigates how housing prices influence the skill composition of migrants in China.

6.3.3. Housing price and migration in China

Before 1978, there was nearly no voluntary population migration in China because the household registration (*hukou*) system, a household-based population management system, strictly controlled the population flow between regions. After the reform and opening-up, the government began to gradually liberalize the free movement of the population but did not change the *hukou* system itself. Generally speaking, residents' *hukou* locations are determined by their place of birth or the location of their parents' *hukou*, which, in turn, directly determines where residents receive their social benefits, including education, medical insurance, and social security. Even with such restrictions, a large-scale population migration emerged since the 1980s, that is, the "great migration." Chan (2012) documented that an estimated 200-250 million rural residents moved to cities and towns within China, while eastern coastal cities were the most attractive.

When massive migration stimulates economic development in China, people's income levels are also rising, thereby stimulating the growth of housing prices. However, this income growth is asymmetric for different groups, given that the national income gap has been gradually widening since the mid-1980s (Luo et al., 2020; Sicular et al., 2007). Zhang et al. (2016) argued that the higher income growth rate for high-income groups has pushed up the equilibrium housing price, thereby increasing the housing cost burden for middle- and low-income groups. Li et al. (2020) and Liu et al. (2017) have documented that low-income/low-skilled households and migrants face serious housing unaffordability predicaments in some megacities.

As a result, housing unaffordability has become another barrier that restricts population migration in addition to the *hukou* system, especially for low-skilled migrants. Scholars have identified that due to housing unaffordability, the attractiveness of megacities for highly educated individuals is gradually decreasing, whereas next-tier cities are becoming more appealing (Chen et al., 2019; Lin et al., 2021). Song and Zhang (2020) found an inverted-U relationship between the housing purchase intention and city size among rural-urban migrants, which means that migrants prefer medium-sized cities to megacities. However, they focused on a specific migrant group and neglected the impact of housing prices on the skill composition of migrants in China, which is the research gap this research fills.

6.4. Theoretical analysis

To illustrate the heterogeneous effects of housing costs on the migration of different skill levels, this subsection introduces housing costs into the self-selection framework within which migrants have heterogeneous migration costs. The model considers two-region migration from region *0* to *1* and assume a group of residents earning w_0 (wage level) in region *0*. When they move to region *1*, they earn a different wage level, w_1. Their wage distributions can be written as follows:

$$\ln w_0 = \mu_0 + \eta_0 s \tag{6.1a}$$

$$\ln w_1 = \mu_1 + \eta_1 s, \tag{6.1b}$$

where w_0 and w_1 represent the wage level of the same residents in the two regions; μ_0 and μ_1 denote the fundamental wages of non-skilled residents; s denotes the skill level of the residents; η_0 and η_1 are the returns to skills in the two regions. Based on the most common migration scenario in China, where migrants move from inland less-developed areas to developed coastal areas, the model assumes that the fundamental wage is higher and returns to skills are lower in the destination regions than in the original regions ($\mu_1 > \mu_0, \eta_1 < \eta_0$)[7].

In addition to wage differentials, residents migrating between two regions will bear certain costs (monetary, such as difference in living costs, or non-monetary, such as difference in language or culture), *i.e.*, migration costs. Migration costs is denoted as C and migration costs in time-equivalent units as $\pi = C/w_0$. Residents will migrate only if their wage differentials between two regions are larger than zero after netting migration costs. The migration condition is shown below

$$I = \ln w_1 - \ln(w_0 + C) \cong \ln w_1 - \ln w_0 - \pi > 0. \tag{6.2}$$

As discussed previously, migration costs will also vary with migrants' skill levels because migrants have different economic bases, adaptive skills, or skills to get information. Therefore,

---

[7] Skilled individuals are highly scarce in inland regions, so the model assumes that returns to skills is higher in these original regions.

the model considers heterogeneous migration costs instead of constant ones following Chiquiar and Hanson (2005) and McKenzie and Rapoport (2010), so

$$\ln \pi = \mu_\pi - \gamma_1 s + \gamma_2 h, \tag{6.3}$$

where the time-equivalent migration costs (π) comprise of three components, the constant migration cost (μ_π), skill-related cost ($\gamma_1 s$), and the housing cost ($\gamma_2 h$). The constant migration cost (μ_π) implies that each migration will incur a fixed migration cost, such as transportation costs. In the meantime, there are other migration costs that vary with migrants' skill levels. The model assumes that these migration costs decrease with the migrant's skills, i.e., $\gamma_1 > 0$. The notion is that high-skilled migrants tend to have better communication, paperwork processing, and information acquisition skills, so they bear fewer migration costs. Finally, migrants bear the housing costs change after migration, which increases the total migration costs, i.e., $\gamma_2 > 0$. As discussed in Section 3, housing costs exert significant negative effects on migrants, especially the newcomers. Since housing costs expenditure occupies a large part of family income, higher housing prices/rents can largely influence their everyday life and induce sizable pecuniary and psychic costs. Therefore, higher housing costs lead to higher migration costs. Then, the following part discusses the migration intention (I) across different skill levels (s).

Figure 6.3 draws (net) wage profiles for individuals staying at the origin (see the solid straight line) and migrants moving to the destination (other lines). Firstly, let us consider a scenario without housing costs ($h = 0$). For non-skilled migrants ($s = 0$), the minimum wage gains from migration cannot cover fundamental migration costs because of poverty constraints (Chiquiar & Hanson, 2005), i.e., $\mu_1 - e^{\mu_\pi} < \mu_0$. For skilled migrants ($s > 0$), their (net) wage profile at the destination is $A = \mu_1 + \eta_1 s - e^{\mu_\pi - \gamma_1 s}$ (see the solid curved line in Figure 6.3). Because $\frac{\partial A}{\partial s} > 0$, the wage profile increases with the migrant's skill level. As shown in Figure 6.3, based on the intersection of wage profiles of stayers and migrants, we get two skill thresholds (s_L, s_U) and directly identify those individuals that will move from region *0* to *1*. Individuals with skill levels less than s_L face overwhelming migration costs and tend to stay, whereas individuals with skill levels higher than s_U can enjoy sufficient return to skills offsetting migration incentives at the origin and tend to stay. Therefore, only individuals with skill levels between s_L and s_U choose to migrate.

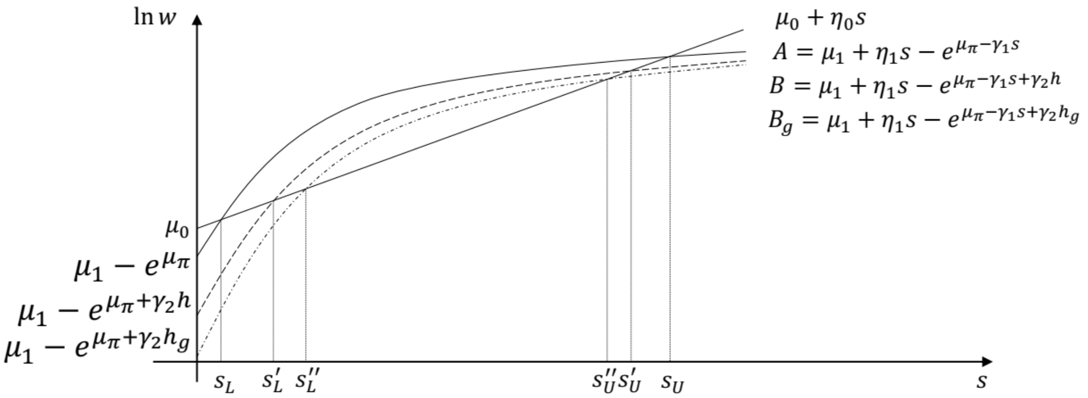

Figure 6.3 Housing costs and self-selection of migrants

Then, another scenario with housing costs ($h > 0$) is considered. As the emergence of housing costs will increase the migration costs of all residents, the burden of housing costs discourages migration despite skill levels (See Appendix A for the mathematical proof). Therefore, the increase in housing prices shifts the wage profile of skilled migrants at the destination downward, which results in a new wage profile, $B = \mu_1 + \eta_1 s - e^{\mu_\pi - \gamma_1 s + \gamma_2 h}$ (the dashed curved line in Figure 6.3).

When housing costs increase, it leads to a new wage profile and two new skill thresholds, s_L' and s_U', where $s_L' > s_L$ and $s_U' < s_U$. As shown in Figure 6.3, higher housing costs discourage both low-skilled (s_L to s_L') and high-skilled (s_U' to s_U) migrants from migration, while sizes of $|s_L - s_L'|$ and $|s_U - s_U'|$ determines the number of discouraged low-skilled and high-skilled migrants, that is, the selection pattern of migrants. According to the model, $|s_L - s_L'|$ is bigger than $|s_U - s_U'|$, i.e., higher housing costs at the destination will have a more significant adverse effect on low-skilled migrants and induce positive selection (see Appendix A for the mathematical proof). Therefore, this research proposes the following hypothesis to test:

H6.1: **High housing costs at the destination discourage the migration of low-skilled migrants more significantly than high-skilled migrants, i.e., a position selection.**

H6.1 proposes the general condition of all migrants. Now, the following part puts more emphasis on the fact that housing costs are not only directly affected by housing prices but also changes with different migrant groups (such as different gender, age cohorts, and employment sectors). As such,

$$h = H(g, p) \tag{6.4}$$

where g denotes different migrant groups; p denotes housing prices. When p keeps unchanged, migrants suffering higher housing costs will have a lower wage profile $B_g = \mu_1 +$

$\eta_1 s - e^{\mu_\pi - \gamma_1 s + \gamma_2 h_g}$ (the dotted curved line in Figure 6.3), which leads to new thresholds s_L'' and s_U''. We have $\frac{|s_L - s_L''| - |s_U - s_U''|}{|s_L - s_L''|} > \frac{|s_L - s_L'| - |s_U - s_U'|}{|s_L - s_L'|}$ because $\frac{\partial s_L/\partial h}{|\partial s_U/\partial h|} > 1$ (see the proof in Appendix A), which implies that a larger proportion of low-skilled groups are crowded out when housing costs increase. In other words, the positive selection degree will be more severe for migrants suffering higher housing costs. Accordingly, this research proposes the second hypothesis to test:

H6.2: The selection degrees are heterogeneous among different migrant groups, especially more severe for migrants with higher housing costs.

The notion behind H6.2 is intuitive. First, migrants with different demographic characteristics, such as age and gender, may react differently to housing unaffordability. Secondly, the settlement intention makes a big difference among migrants. Liu et al. (2017) revealed that migrants who tend to settle down strive to get formal housing. Therefore, these individuals bear higher housing costs in their life course and should be more sensitive to housing prices. Thirdly, migrants in different employment sectors face different work conditions and labor welfares. For example, state-owned enterprises will provide higher provident funds for house purchasing or directly provide housing after meeting certain conditions, which are rare in private enterprises.

The theoretical analysis only provides a simple simulation of reality, which requires data validation. The following content conducts an empirical study to verify these two hypotheses.

6.5. Data and methodology

6.5.1. Empirical methodology

This research aims to investigate how housing prices influence the skill level of migrants. As discussed in the theoretical analysis, multiple factors will influence individuals' migration decisions. However, only skill-related factors (heterogeneous migration costs) will influence migrants' skill level because others that do not vary with skills (constant migration costs) will be offset. Accordingly, this research proposes the following Equation to estimate the effects of housing prices on migrants' skill level:

$$Skill_{pij} = \alpha_0 + \alpha_1\left(H_j - H_i\right) + \alpha_2 D_p + \alpha_3 X_i + \alpha_4 X_j + \theta_i + \theta_j + \varepsilon_{pij}, \qquad (6.5)$$

where $Skill_{pij}$ denotes the skill level of migrant p who has moved from city i to city j; H denotes the urban housing cost indicators; D denotes migrants' demographic characteristics; X denotes city-level influencing factors; θ controls province-level fixed effects of the original and destination cities. ε_{pij} is the standard error.

This research takes education level as migrants' skill level ($Skill_{pij}$) as it is the most essential and measurable skill and is widely used in the literature (Jaschke & Keita, 2021; Lucas, 1988; McKenzie & Rapoport, 2010). In detail, this research takes two measures to represent personal education level: years of schooling attended and obtainment of a college degree (or above). Years of schooling can directly capture the length of learning time and is a good proxy for human capital and individuals' skill levels (Abramitzky et al., 2021; McKenzie & Rapoport, 2010; Mincer, 1970, 1974). Obtainment of a college degree (or above) indicates one's skill position as the college degree (or above) is particularly important in today's knowledge economy. Numerous scholars employ it to define skilled and non-skilled migrants in recent migration literature (Beine et al., 2011; Grogger & Hanson, 2011; Shen & Liu, 2016).

H is the mainly concerned variable, which reflects the condition of urban housing markets. According to the theoretical model, higher housing costs at destination cities lead to a positive selection of migrants. Therefore, α_1 is significant and positive. This research uses urban average housing prices to capture housing costs in different cities as it is the most direct housing cost indicator. To ensure the robustness of the results, this research introduces other housing cost indicators which are discussed in Appendix B2.

To capture other influencing factors, the empirical model introduces individual-level and urban-level controls. Individual-level controls include migrants' demography characteristics such as age, gender, marital status, family size, and *hukou* status. In addition, the migration distance, captured by a dummy variable denoting intra-provincial migration, will also influence the skill level as high-skilled migrants tend to migrate further. Urban-level controls include other heterogeneous migration costs. As high-skilled migrants prefer larger, more developed cities, this research employs urban population, GRP per capita, and GRP growth rate to capture these features. Industry structure determines different demands for labor with different skills; thus, this research uses the ratio of the tertiary industry to the secondary industry's GRP percentage to measure it. Finally, urban amenities enjoyed by migrants also vary with their different skills; this research uses education facilities per capita, medical facilities per capita, and green covered area ratio to measure these.

Although having introduced numerous controlling variables and controlled province fixed effects, the model may still suffer from endogenous issues induced by omitted variables. This research considers utilizing the instrumental variable to solve this concern. Land supply is the fundamental of housing markets and largely determines the cost of housing development, thereby directly influencing the housing price (Huang & Tang, 2012). In the meantime, land supply intuitively does not affect migration through other channels. Consequently, land supply is highly suitable to instrument housing prices as it satisfies the requirement of an efficient instrument variable (Liang et al., 2016). Empirically, this research employs the one-year lagged land supply per capita to capture actual land supply conditions since there is a time lag in the response of housing prices to land supply.

6.5.2. Data

This research employs the *China Migrant Dynamic Survey* (CMDS) database to conduct empirical research. CMDS is an annual national household-level survey initiated by the National Health Commission of China; it examines the survival and development status, migration trends, and characteristics of China's floating population. This survey only targets migrants without local *hukou* (floating population), residing in the 32 provinces or provincial municipalities, except Hong Kong, Macao, and Taiwan. It provides detailed information about migrants, including their household members, family finances, housing, employment, migration history, and health.

This research adopts the 2017 wave survey data because only this wave's data provides the original address of migrants at the prefecture city level. Although CMDS records information about the entire household, the research sample only investigates the migration of the head of the household since the head mostly represents the condition of the entire household. Regarding economic migration, the research sample excludes migrants whose reason for migration is not employment or business. The age of migrants is limited to between 15 and 65. Furthermore, CMDS does not only cover newcomers but also those who have migrated for a long time but have not obtained the local *hukou*. However, it does not provide the skill or employment status of migrants before migration. Therefore, the research sample mainly considers newcomers that migrated after 2016 because their education level has not changed drastically and thus can represent the skill level at the time of migration. The research sample also concentrates on inter-city urban migrants and excludes those intra-city migrants or migrants living in rural areas. Finally, there are 15,815 newcomers for the analysis.

This research collects the average housing price at the prefecture city level from the CEIC database. The CEIC is a very authoritative database that organizes and collects data from official regional yearbooks, i.e., the annual average commercial housing selling price. In addition to housing price, the empirical model introduces other urban-level controlling variables from *China City Statistical Yearbook 2017*, which records urban macro statistical data in 2016. Some cities also have problems with missing data in the CEIC database and the *China City Statistical Yearbook 2017*, so not all the migrant samples were used for the regression. Table 6.1 shows the statistical description of data.

Table 6.1 Statistical description of variables

	N	Mean	Sd	Min	Max
Individual-level variables (newcomers)					
Household head's years of schooling	15,815	11.010	3.252	0	19
Household head's obtainment of the college degree	15,815	0.245	0.430	0	1
Household head's age (year old)	15,815	31.710	9.737	15	65
Household head's gender (male = 1)	15,815	0.550	0.498	0	1
Household head's marital status (unmarried = 1)	15,815	0.349	0.477	0	1
Household head's *hukou* status (rural *hukou* = 1)	15,815	0.777	0.416	0	1
Household size (person)	15,815	2.564	1.295	1	10
Settlement decision (intend to live over five years = 1)	15,815	0.699	0.459	0	1
Hukou transfer intention (intend to transfer = 1)	15,815	0.365	0.482	0	1
Intra-provincial migration (intra-provincial = 1)	15,815	0.438	0.496	0	1
City-level variables					
Average housing prices in 2016 (Yuan/m^2)	264	5599.048	4054.577	2517	45146
Average housing prices in 2015 (Yuan/m^2)	262	5290.269	3214.305	2248	33942
Household registered population at year-end (10,000 persons)	291	447.869	322.645	21	3392
Per capita gross regional product (Yuan)	291	53492.930	30987.830	11892	215488
Gross regional product growth rate (%)	291	6.997	3.522	-12.3	12.4
Industry structure (ratio of tertiary industry to second industry's GRP percentage)	291	1.060	0.531	0.370	4.166
Total number of middle and primary schools per 10,000 persons	291	1.803	0.861	0.578	6.382
Total number of hospitals and health centers per 10,000 persons	291	0.594	0.708	0.087	8.929
Green covered area as percentages of completed area (%)	289	39.244	6.507	3.07	61.58

6.6. Empirical results

6.6.1. Housing prices and migrant selection

As shown in Table 6.2, this research estimates the effects of housing prices on migrant education selection using two different education measures. Column (1) of Table 6.2 demonstrates that the difference in housing price between destination and original cities (*Delta ln housing price)* leads to a significant positive selection of migrants by education when controlling individual-level variables. After controlling urban-level variables in Columns (2) of Table 6.2, the coefficient of the housing price difference is still significant and positive. Statistically, a 1% increase in the housing price difference increases 0.297 schooling years of migrants, implying that higher educated migrants are selected into destination cities with higher housing prices, which supports **H6.1**. Similar results are reported in Columns (4) and (5) in Table 6.2 when using obtainment of the college degree as skill levels, which shows that *Delta ln housing price* has significant positive effects on migrants' obtainment of the college degree.

Table 6.2 Housing prices and migrant selection

	Years of schooling attended			Obtainment of the college degree (dummy)		
	(1)	(2)	(3)	(4)	(5)	(6)
Delta ln housing price	0.379***	0.297***	1.299***	0.166***	0.188***	0.463**
	(0.061)	(0.095)	(0. 0.327)	(0.033)	(0.053)	(0.182)
Age	-0.119***	-0.119***	-0.118***	-0.040***	-0.041***	-0.040***
	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)
Gender (male)	0.142***	0.160***	0.161***	-0.069***	-0.061**	-0.061**
	(0.048)	(0.047)	(0.048)	(0.026)	(0.027)	(0.027)
Marital status (unmarried)	-0.311***	-0.279***	-0.283***	-0.120***	-0.101**	-0.102**
	(0.076)	(0.076)	(0.076)	(0.041)	(0.042)	(0.042)
Hukou status (rural)	-2.237***	-2.143***	-2.182***	-0.951***	-0.916***	-0.925***

	(0.059)	(0.059)	(0.060)	(0.031)	(0.031)	(0.031)
Family size	-0.271***	-0.249***	-0.257***	-0.137***	-0.128***	-0.130***
	(0.025)	(0.025)	(0.025)	(0.015)	(0.015)	(0.015)
Migration scale (intra-provincial)	0.388***	0.369***	0.339***	0.206***	0.199***	0.190***
	(0.066)	(0.066)	(0.067)	(0.036)	(0.037)	(0.037)
Ori ln population		0.110*	0.290***		0.055	0.104**
		(0.061)	(0.083)		(0.034)	(0.046)
Des ln population		0.066	0.099		0.053	0.061
		(0.076)	(0.077)		(0.046)	(0.046)
Ori ln GRP		0.736***	1.129***		0.369***	0.475***
		(0.085)	(0.149)		(0.047)	(0.081)
Des ln GRP		0.156	-0.454**		0.014	-0.157
		(0.105)	(0.217)		(0.062)	(0.124)
Ori GRP growth rate		-0.013	0.004		0.001	0.006
		(0.020)	(0.021)		(0.011)	(0.012)
Des GRP growth rate		0.016	-0.004		-0.006	-0.011
		(0.0.019)	(0.020)		(0.011)	(0.011)
Ori industry structure		0.268***	0.436***		0.133**	0.178***
		(0.095)	(0.109)		(0.052)	(0.059)
Des industry structure		0.560***	0.088**		0.257***	0.128
		(0.100)	(0.178)		(0.058)	(0.100)
Ori School number per capita		-0.102*	-0.085		-0.040	-0.036
		(0.058)	(0.058)		(0.033)	(0.033)
Des School number per capita		-0.175**	-0.270***		-0.131***	-0.157***
		(0.016)	(0.079)		(0.044)	(0.047)
Ori Hospital number per capita		0.165**	0.169**		0.037	0.037
		(0.065)	(0.065)		(0.034)	(0.034)
Des Hospital number per capita		-0.338***	-0.352**		-0.121	-0.126
		(0.099)	(0.100)		(0.084)	(0.084)
Ori Green area rate		0.009*	0.013**		0.005	0.006*
		(0.005)	(0.005)		(0.003)	(0.003)
Des Green area rate		-0.011	-0.026***		-0.004	-0.008
		(0.008)	(0.009)		(0.005)	(0.005)
Constant	21.370***	6.580***	9.717***	3.307***	-3.365***	-2.932***
	(0.698)	(1.907)	(2.145)	(0.398)	(1.085)	(1.165)
Origin FE	Yes	Yes	Yes	Yes	Yes	Yes
Destination FE	Yes	Yes	Yes	Yes	Yes	Yes
First stage: Delta ln lagged land supply			-0.083***			-0.083***
			(0.002)			(0.002)
Method	OLS	OLS	2SLS+IV	Probit	Probit	Probit+IV
Observations	13,206	13,206	13,206	13,206	13,206	13,206
Adjusted/Pseudo R^2	0.288	0.300	0.298	0.178	0.190	

Notes: 1. *Delta* means the differences in the destination and original cities; *Ori* and *Des* refer to the variables of the original and destination cities, respectively. 2. ***, **, and * denote significance at the 1, 5, and 10 percent levels, respectively. 3. The standard errors are in parentheses.

To ensure the robustness of empirical results, this research compares different estimation methods. The instrumental variable is utilized in the empirical model to solve the endogeneity concern, as shown in Columns (3) and (6) in Table 6.2. The first stage results show that *Delta ln lagged land supply* is an efficient instrumental variable for *Delta ln housing price*. 2SLS regression results show that *Delta ln housing price* still has significant positive effects on migrants' skill levels, ensuring the robustness of the hypothesis after solving the endogenous issue.

Migrants' housing tenure choices reveal their entry to different housing markets in destination cities, which may influence the selection outcomes. As shown in Columns (1) and (2) of Table 6.3, this research compares migrants living in unit/employer-supplied, rental, self-purchased, public, and other housing. The results show that migrants living in unit/employer-supplied houses are less positively (or even not) selected by housing prices given the insignificant coefficients. The possible reason is that these migrants do not need to enter the private housing market because of cheaper/free housing supplied by their employers, at least at the early stage of migration. By contrast, given the significant coefficients of interactions, migrants living in rental or self-purchased housing are more sensitive to the private housing market since they bear higher housing costs, especially those living in self-purchased housing.

Table 6.3 Housing prices and migrant selection: different housing tenures

	Years of schooling attended	Obtainment of the college degree (dummy)

	(1)	(2)
Delta ln housing price	-0.071	-0.048
	(0.122)	(0.072)
Housing tenures: (Taking Unit/Employer-supplied housing as the reference)		
Rental housing	0.345***	0.117**
	(0.087)	(0.051)
Self-purchased housing	1.443***	0.679***
	(0.118)	(0.066)
Public housing	0.522**	0.260*
	(0.256)	(0.141)
Others	0.127	0.015
	(0.244)	(0.150)
*Delta ln housing price * Housing tenures: (Taking Unit/Employer-supplied housing as the reference)*		
Rental housing	0.403***	0.273***
	(0.093)	(0.055)
Self-purchased housing	0.744***	0.408***
	(0.150)	(0.082)
Public housing	-0.196	-0.153
	(0.471)	(0.259)
Others	0.758 ***	0.365**
	(0.278)	(0.160)
Individual-level controls	Yes	Yes
Urban-level controls	Yes	Yes
Origin FE	Yes	Yes
Destination FE	Yes	Yes
Method	OLS	Probit
Observations	13206	13206
Adj R^2	0.323	0.214

Notes: 1. *Delta* means the differences in the destination and original cities; *Ori* and *Des* refer to the variables of the original and destination cities, respectively. 2. ***, **, and * denote significance at the 1, 5, and 10 percent levels, respectively. 3. The standard errors are in parentheses.

6.6.2. Housing prices and selection of different migrant groups

This subsection further compares selection degrees among different migrant groups to test H6.2. In detail, this research introduces different interaction items of housing prices and migrant groups into Equation (6.5). As shown in Table 6.4, this research considers different demographic characteristics of household heads. Columns (1) and (5) of Table 6.4 shows that housing prices more positively select young migrants. Compared with old migrants, young migrants are at the beginning of their careers. They do not have enough savings to cover huge housing expenditures, so highly educated migrants are more likely to find jobs with sufficient salaries to survive in cities with high housing prices. In contrast, the difference is not significant between different genders and *hukou* status.

Table 6.4 Housing prices and migrant selection: different demographic characteristics

	Years of schooling attended			Obtainment of the college degree (dummy)		
	(1)	(2)	(3)	(4)	(5)	(6)
Delta ln housing price	-0.248**	0.357***	0.354***	-0.204***	0.213***	0.177***
	(0.114)	(0.103)	(0.117)	(0.071)	(0.057)	(0.062)
Delta ln housing price	0.667***			0.440***		
** Young (15~35)*	(0.077)			(0.052)		
Delta ln housing price		-0.110			-0.049	
** Male*		(0.075)			(0.041)	
Delta ln housing price			-0.074			0.015
** Rural hukou*			(0.089)			(0.046)
Individual-level controls	Yes	Yes	Yes	Yes	Yes	Yes
Urban-level controls	Yes	Yes	Yes	Yes	Yes	Yes
Origin FE	Yes	Yes	Yes	Yes	Yes	Yes
Destination FE	Yes	Yes	Yes	Yes	Yes	Yes
Method	OLS	OLS	OLS	Probit	Probit	Probit
Observations	13206	13206	13206	13206	13206	13206
Adj R^2	0.304	0.300	0.300	0.195	0.300	0.300

Notes: 1. Delta means the differences in the destination and original cities; Ori and Des refer to the variables of the original and destination cities, respectively. 2. Individual-level controls have included variables of age, gender, and *hukou* status. 3. ***, **, and * denote significance at the 1, 5, and 10 percent levels, respectively. 4. The standard errors are in parentheses.

Then, this research compares the selection degrees among migrants with different settlement intentions. Columns (1) of Table 6.5 shows that a 1% increase in the housing price difference increases the 0.372 schooling years of migrants who intend to live at the destination over five years, implying that they are more significantly selected by housing prices than their counterparts. As shown in Columns (2) and (4) of Table 6.5, this research also considers migrants' *hukou* transfer intention and find similar results that migrants who intend to transfer their *hukou* are more positively selected by housing prices. These results are highly consistent with the literature on return migration that failed migrants will return home and accentuate the original selection pattern of migration (Borjas & Bratsberg, 1996; Wang & Fan, 2006; Zhang et al., 2020).

Table 6.5 Housing prices and migrant selection: different settlement intentions

	Years of schooling attended		Obtainment of the college degree (dummy)	
	(1)	(2)	(3)	(4)
Delta ln housing price	0.143	0.170*	0.126**	0.135**
	(0.111)	(0.100)	(0.062)	(0.057)
Settle down	0.132*		0.110**	
	(0.075)		(0.043)	
*Delta ln housing price * Settle down*	0.229***		0.091**	
	(0.081)		(0.045)	
Hukou transfer		0.530***		0.266***
		(0.073)		(0.040)
*Delta ln housing price * Hukou transfer*		0.200**		0.081*
		(0.079)		(0.043)
Individual-level controls	Yes	Yes	Yes	Yes
Urban-level controls	Yes	Yes	Yes	Yes
Origin FE	Yes	Yes	Yes	Yes
Destination FE	Yes	Yes	Yes	Yes
Method	OLS	OLS	Probit	Probit
Observations	13206	13206	13206	13206
Adj R^2	0.302	0.310	0.192	0.195

Notes: 1. Delta means the differences in the destination and original cities; Ori and Des refer to the variables of the original and destination cities, respectively. 2. ∗∗∗, ∗∗, and ∗ denote significance at the 1, 5, and 10 percent levels, respectively. 3. The standard errors are in parentheses.

Finally, this research estimates the heterogeneous selection degrees of migrants in different types of enterprises. Columns (1) and (3) in Table 6.6 show that positive education selection induced by housing price is more significant among migrants in private enterprises than state-owned enterprises. The possible reason is that state-owned enterprises provide better welfare, including free-supplied housing, housing subsidies, and provident funds. These benefits lighten their housing burden in cities with high housing prices; thus, they are less positively selected by housing prices. To eliminate the influence of Unit/Employer-supplied housing, this research excludes migrants who live in Unit/Employer-supplied housing and estimate the model based on the sub-sample. Columns (2) and (4) of Table 6.6 show similar results to the entire sample's results, ensuring the reliability of results.

Table 6.6 Housing prices and migrant selection: different Enterprise types

	Years of schooling attended		Obtainment of the college degree (dummy)	
	(1) Full sample	(2) Sub-sample	(3) Full sample	(4) Sub-sample
Delta ln housing price	-0.162 (0.159)	-0.456** (0.190)	-0.067 (0.086)	-0.208* (0.107)
Enterprise types: (Taking Enterprise types one as the reference)				
Type two	-1.722*** (0.121)	-2.181*** (0.143)	-0.923*** (0.065)	-1.149*** (0.082)
Type three	-1.274*** (0.218)	-1.903*** (0.273)	-0.686*** (0.117)	-1.017*** (0.148)
*Delta ln housing price * Enterprise types: (Taking Enterprise types one as the reference)*				
Type two	0.447*** (0.133)	0.626*** (0.164)	0.246*** (0.071)	0.317*** (0.092)
Type three	0.184 (0.210)	0.604** (0.253)	0.092 (0.112)	0.309** (0.136)
Individual-level controls	Yes	Yes	Yes	Yes
Urban-level controls	Yes	Yes	Yes	Yes

Origin FE	Yes	Yes	Yes	Yes
Destination FE	Yes	Yes	Yes	Yes
Method	OLS	OLS	Probit	Probit
Observations	11417	8472	11417	8471
Adj R^2	0.302	0.325	0.201	0.219

Notes: 1. *Enterprise type one* includes official institutions, state-owned enterprises, and collective enterprises; *Enterprise type two* includes stock/associated enterprises, private enterprises, and self-employed; *Enterprise type three* includes wholly foreign-owned enterprises and Chinese-foreign equity joint ventures. 2. Columns (2) and (4) exclude migrants who live in *Unit/Employer-supplied housing*. 2. *Delta* means the difference in the variables between the destination and original cities; *Ori* and *Des* refer to variables of the original and destination cities, respectively. 3. ***, **, and * denote significance at the 1, 5, and 10 percent levels, respectively. 4. The standard errors are in parentheses.

6.7. Discussions

This research mainly discusses the effects of average housing prices on migrants' skill levels, thereby missing some crucial housing dimensions in China. For most low-educated migrants, an urban village is suitable for them to stay first after migration because of its relatively low rent and high accessibility to employment locations (Hui et al., 2014; Tao et al., 2014). However, this choice is gradually disappearing in big cities and being replaced by gentrified housing in recent years, leaving migrants less room to survive in these cities. In Hangzhou, housing rent has caused a considerable expenditure burden; 14.1% of migrants plan to return home after the local government demolishes urban villages (Zeng et al., 2019). In this regard, the positive selection of migrants by housing price is becoming more and more severe.

Apart from the private housing market, the affordable housing system is also unfriendly to migrants. Shi et al. (2016) summarized various types of public housing and found that most public housing types require local *hukou*. In contrast, public rental housing is the only choice for migrants with slightly lower rent than the private market. According to a report of the *China Household Family Survey* (CHFS)[8], in 2019, only 2.1% of migrants lived in public housing, whereas the proportion of the locals was 18%. Furthermore, there is a mismatch between public housing and income groups. Some governments use public housing as a tool to attract highly educated talents to facilitate local economic development, overlooking the more urgent needs of low-educated/low-income migrants. The discriminatory housing policies have intensified the education selection of migrants induced by private housing markets, squeezing low-educated migrant groups out.

As a result, low-educated migrants regard destination cities as places to earn money instead of homes to live in (Zheng et al., 2009). Results also partially support the assertion that failed migrants return home, reinforcing the positive selection of permanent migrants by housing prices. Unlike low-educated counterparts, highly educated migrants not only have higher incomes but also have better access to obtain local *hukou* and enjoy local social welfare. Interestingly, mid-educated migrants are trapped in an awkward position. Their income cannot cover housing expenditure in destination cities, whereas there may be no suitable job that matches their education level in home cities. Consequently, more and more migrants flee to second-, third-, and fourth-tier cities with relatively lower migration costs (Chen et al., 2010;

[8] Please see the report "Paying attention to new citizens is the future direction of housing security (关注新市民是住房保障未来方向)". *https://chfs.swufe.edu.cn/thinktank/columnarticle.html?id=2367*

Lin et al., 2021; Song & Zhang, 2020).

6.8. Conclusions

Compared with international migration, housing cost plays a crucial role in decisions in internal migration since it significantly influences the overall migration cost. Previous studies have identified its negative effects on migration. However, little is known about its effects on the skill composition of migrants, i.e., how migrants are self-selected. To fill this research gap and achieve Objective 2.4, this chapter introduces housing costs into the self-selection model and develops a theoretical framework. This framework proposes the following two hypotheses: i) unaffordable housing prices at the destination discourage the migration of low-skilled migrants more significantly than high-skilled migrants (positive selection), and ii) this positive selection varies in migrant groups. Using China's internal migration as a case study, this chapter further develops an empirical model to test these hypotheses by employing the 2017 wave of CMDS data, the *CEIC* database, and the *China City Statistical Yearbook*. The major findings are as follows.

First, the results support H6.1, indicating that a 1% increase in the housing price difference increases 0.297 schooling years of migrants. This implies that higher housing price difference between the original and destination cities leads to a positive education selection of migrants. Therefore, high housing prices in destination cities have raised migration costs and caused a bigger shock to low-educated migrants, discouraging their migration. Second, results show that there are significant heterogeneous effects between different migrant groups, supporting H6.2. The education selection is more significant among young migrants who live in rental or self-purchased houses and work in private enterprises. Also notable is the severer positive selection of migrants who intends to settle in the migrant cities.

The results provide several urban policy implications. First, the research has emphasized the uneven conditions of migrants under the influence of housing unaffordability. The government should pay attention to the selective migration caused by housing unaffordability since it may enlarge the disparity in regional development, hence damaging national economic development. Second, low-skilled migrants are also crucial and indispensable to sustainable urbanization and economic growth (Eeckhout et al., 2014). To house these low-skilled migrants, more flexible and low-skilled-oriented housing policies should be promulgated to meet the needs of different migrant groups based on their different situations. In summary, housing low-skilled migrants is always a crucial and urgent issue to new urbanization and urban development.

CHAPTER 7. CONCLUSIONS

7.1. Summary of major findings

7.1.1. The education selection pattern of migrants in China

The most important finding of this thesis is that the inter-regional migration in China reveals a U-shaped selection pattern of migrants regarding education levels. Although highly educated (four-year college degree and above) migrants still have higher migration probabilities than others, the emigration rate does not decrease monotonically with education levels. Instead, mid-educated (senior high school or three-year college degree) migrants surprisingly have the lowest propensity to emigrate. This selection pattern of inter-regional migrants is unique and inconsistent with selection patterns in most cases in international migration, such as positive selection (Grogger & Hanson, 2011), negative selection (Abramitzky et al., 2012; Borjas, 2008), or intermediate selection (Chiquiar & Hanson, 2005; Gould & Moav, 2016). This pattern is also differing from the typical positive selection of labor moving to big cities, found in the US and European countries (Behrens et al., 2014; De la Roca, 2017). This difference raises the critical question of what specific mechanisms in China have induced the U-shaped pattern. This thesis proposes four new impact channels (some are unique to China and some are not) attributing to this unique pattern behind return to skills and heterogeneous migration costs.

7.1.2. Underlying factor one: the *hukou* system

The *hukou* system is perhaps most responsible for this selection pattern. Typical selective migration policies influence migration through directly restricting (a certain part of) migrants (Beine et al., 2011; Clark et al., 2007; Mayda, 2010). In contrast, this research proposes another direction of how selective migration policies influence the migrant selection pattern through indirect channels. The *hukou* system does not directly prohibit the free movement of migrants but indirectly influences their economic returns and living costs in destination regions.

The proposed new theoretical framework in Chapter 3 suggests that, on the one hand, this system will restrict *hukou*-related social amenities for migrants without local *hukou*; on the other hand, these migrants will simultaneously face labor market discrimination (which influences return to skills) that increases with skill levels. Also, the local *hukou* obtainment ability increases with skill levels because local governments prefer high-skilled labor. With the combined effect of these channels, mid-skilled migrants obtain the lowest economic migration returns from inter-regional migration, resulting in the lowest migration probabilities. Empirical results in Chapter 3 also verify these inferences. As such, the *hukou* system has reshaped the monotonically positive selection pattern by restricting *hukou*-related social amenities and distorting return to skills by wage discrimination.

7.1.3. Underlying factor two: different income inequality components

The lower degree of income inequality in coastal regions may explain the high emigration probabilities of low-skilled migrants from inland regions. In the past few decades, instead of high-skilled labor, the economic development in coastal regions seems to demand more low-skilled labor, thus boosting their income levels relative to that of high-skilled ones. As a result, chapter 4 demonstrates the lower degrees of income inequality (return to skills) in coastal regions relative to inland regions. The empirical model has confirmed that this inequality has resulted in the higher migration possibilities of low-skilled migrants from inland to coastal regions.

Furthermore, in addition to the income inequality component induced by labor market supply and demand, inequality of opportunity, induced by uneven social opportunities, similarly leads to significant positive selection of migrants through indirect channels. These results imply that these low-educated migrants from inland to coastal regions are not only fleeing poverty but also chasing more social opportunities, which, to a large extent, determines both their own and their next-generation outcomes. In contrast, high inequality of opportunity in inland regions may retain high-skilled individuals and mitigate the brain drain but damage the initial human capital formation.

On the one hand, results regarding inequality of opportunity provide a new angle to explain the wave of migrant workers in China, in addition to direct wage returns. Migration may likewise be driven by the uneven distribution of social opportunities. These social opportunities ultimately affect migration returns indirectly in the form of income. On the other hand, this research proposes another impact channel of inequality of opportunity that influences regional human capital accumulation, i.e., population migration. Previous studies have emphasized that high inequality of opportunity may discourage the investment in the human capital of children and thus undermines human capital formation. This research found that under this inequality condition, migration may provide a good way to improve one's income and create a better environment for offspring in the meantime.

7.1.4. Underlying factor three: technological change

However, the high demand trend for low-skilled labor is changing, though. This is due to the economic transition and technological evolution in recent years, which stimulates the demand for high-skilled labor and increases returns to skills and degrees of income inequality. In other words, technological change will lead to a positive selection of migrants to developed regions. Chapter 5 provides a persuasive case regarding how the installation of industrial robots displaces and reinstates labor, which influences the occupation and skill selection of migrants. Empirical results reveal that although robot exposure has led to more migrants in the production sector, these migrants are mainly high-skilled ones. Low-skilled migrants have been displaced and crowded out from the local production sector. These results herald the continuing decline in the importance of cheap and low-skilled labor to the development of high-tech industries. As a result, higher migration probabilities of low-skilled individuals, documented in Chapters 3 and 4, will come to an end in the near future.

Furthermore, our empirical results found low-skilled migrants bear the most severe crowding-out effect, which is inconsistent with the theory proposed by Autor et al. (2003) that automation will replace mid-skilled labor and disproportionately lead to more labor at the two extremes of the skill distribution. The gap in industrial development between China and developed countries seems to explain this inconsistency. The manufacturing industry in China is still in the period of transition from labor-intensive mode, while low-skilled workers consist of the dominant workforce. The installation of industrial robots is still displacing this part of labor. With the development of the Chinese industry, we can witness that an increasing number of highly skilled workers are gradually being threatened by automation.

7.1.5. Underlying factor four: housing costs

Although undertaking lower transportation and psychological migration costs, this research proposes that internal migrants still face considerable costs in the destination. In addition to the *hukou* system limiting social benefits, migrants are simultaneously facing unaffordable housing. Similar to technological change, soaring housing prices in big cities also lead to a positive selection of migrants to these cities. Consistent with the proposed theoretical model, empirical results show that an 1% increase in the housing price difference increases 0.297 schooling years of migrants. This confirms that housing prices lead to a positive selection of migrants selecting high-skilled migrants and crowding out low-skilled ones. High housing prices have induced significant housing costs for low-skilled migrants, thus discouraging their migration. This crowding-out effect is particularly significant among young migrants who live in rental or self-purchased houses and work in private enterprises. As a result, only those mid- and high-skilled migrants who can afford the high housing costs will choose to settle down in the destination cities, leading to a severer positive selection among migrants who intend to settle.

In summary, the underlying mechanism of the U-shaped selection of migrants in China is a complex interweaving of multiple factors. The inherent *hukou* system has built a huge invisible selective mechanism for migrants through restricting social amenities and distorting return to skills, which results in the worst situation for mid-skilled migrants. In contrast, low-skilled migrants enjoy lower degrees of income inequality to compensate for their income levels and thus have higher migration probabilities. However, this pattern is gradually fading away because of economic development and technological changes, which create higher demand and income levels for high-skilled migrants. The resulting sky-rocking housing prices of these two factors further accelerate this progress and crowd out low-skilled migrants from unaffordable cities.

7.2. Policy implications from major findings

This research has identified a U-shaped migrant selection pattern, which demonstrates heterogeneous conditions faced by migrants with varying skill levels. High-skilled individuals inherently have higher mobility given the lower migration costs they face. This research found that the *hukou* system and unaffordable housing costs have reinforced their advantages over other migrant groups. Local governments take advantage of the *hukou* system for attracting preferred high-skilled migrants by granting them with local *hukou* identities. Bedsides, “talent introduction policies” have been launched in various cities to attract target high-skilled groups in recent years. There is no doubt that these tools can promote regional human capital accumulation and economic development, especially in the new era of the knowledge economy.

However, highly skilled individuals may still choose not to migrate due to the high level of inequality of opportunity in less developed original areas. Chapter 4 reveals that a high level of IOP will retain high-educated individuals. This mechanism may induce the misallocation of human capital since it has distorted labor allocation across regions in a free labor market, further resulting in human capital waste and the loss of economic growth at the national level. With technological advances and industrial upgrading, the demand for knowledge and high skills will become more intense. For example, Chapter 5 has shown the high demand for high-skilled labor in the process of automation. As a result, the resulting consequences of the human capital misallocation induced by IOP will be more intense. The government should adopt social policies to increase economic returns to personal efforts rather than endowed social opportunities so as to reduce inequality of opportunity, optimize the matching of skill and job, and promote society's harmonious development.

Contrary to the high-skilled migrants, the condition for low-skilled migrants is getting worse, although they have once enjoyed the huge dividends brought by economic development in the past decades. On the one hand, the *hukou* system has excluded them from the local welfare system, undermining their living conditions. This research suggests that the *hukou* system has been transformed into a selection system allocating migrants to different levels of cities according to their skill levels, creating an enormous spatial stratified structure. Similar to IOP, this semi-mandatory skill matching may distort the free match of labor skills and markets, which, in turn, reduces productivity, damages economic development, and induces regional inequalities. On the other hand, technological change, as well as the resulting industrial upgrading, have gradually used capital (such as machines) to replace low-skilled jobs, which lowers the demand for low-skilled labor and damages their wages. In such cases, these replaced migrants either go home or continue to flow into the service sector in big cities. Yet, the rising housing prices in big cities have raised high and asymmetrical migration costs to low-skilled migrants and thus undermined their real wages. As a result, both options mean worse work conditions and income levels.

This poor condition for low-skilled migrants will bring some severe consequences to society. On the one hand, these low-skilled labor are imperative to urban economic performance, given their complementarities for high-skilled labor (Eeckhout et al., 2014). The outflow of low-skilled migrants from big cities will inevitably affect the work efficiency of other groups, as daily life is no longer effectively supported. On the other hand, the unfair condition for low-skilled migrants may further result in severe social inequality issues and affect social harmony.

Therefore, there is an urgent need to consider how to settle these low-skilled migrants. There are three potential directions for policy development that can be considered to mitigate or address these issues according to this thesis. First, a unified national welfare system is strongly demanded in China to improve the living condition of low-skilled migrants. Even after several rounds of *hukou* reform in the past twenty years, the current system is still skill-preferred and unfriendly to low-skilled groups. Second, workers facing technological change should be granted a path to a new-style job at the current company or a skill that can direct them to a new company. Third, more flexible and low-skilled-oriented housing policies should be promulgated to meet the needs of different migrant groups based on their different situations. By doing so, these low-skilled groups are able to regain their role in the destination cities, rather than being outright abandoned, which can also facilitate social harmony and equality.

Interestingly, mid-skilled migrants are trapped in an awkward position. When facing greater labor market discrimination than other groups due to the *hukou* system and bearing the greatest loss of economic gains as a result, they must also confront unaffordable housing costs similar to those of the low-skilled. Consequently, they may obtain the lowest economic return from inter-regional migration, consistent with their lowest migration probabilities. This situation may be more significant along with the automation process in China in the near future, given that machines most threaten mid-skilled labor with routine-type jobs according to the theory proposed by Autor et al. (2003). To escape this trap, more and more migrants flee to second-, third-, and fourth-tier cities with relatively lower migration costs and higher real wages (Chen et al., 2010; Lin et al., 2021; Song & Zhang, 2020). Considering that they are still the main force of industrial development at the current stage of development, the government and society should pay more attention to this particular group and introduce more policies and measures to improve their living conditions so as to play a greater role in destination cities.

7.3. Future research directions

This research has comprehensively investigated the underlying mechanism of self-selection in China but paid no attention to the consequences of migrant selection. In recent years, studies have begun to emphasize the critical role of selective migration in regional development and propose that positive selection of migrants may enlarge the regional development gaps (Duranton & Puga, 2004; Kanbur & Rapoport, 2005; Puga, 2010; Van Oort, 2007). Unfortunately, these discussions are pretty few in China. China has a large population, a vast territory, and a severe imbalance in regional development. As a result, how the skill composition of selective migration influences regional disparity beyond the migration scale is imperative in contemporary society.

In the meantime, selective migration may also influence the within-regional social inequality. The large flow of low-skilled migrants into developed coastal cities will inevitably cause inequality issues in destination cities, which have been heated discussed for a long time (Chen et al., 2018; Hao et al., 2020). However, they usually neglect the skill composition of migrants. Future studies should recheck these critical issues from the perspective of selective migration when the data is available.

Finally, previous studies concerning internal migration have paid limited attention to the consequences to the origin regions. For example, in international migration, Beine et al. (2001, 2008) propose that the positive selection of emigrants to prosperous countries will induce skilled migration prospects on stayers in original countries, which thus facilitates the gross human capital formation. This mechanism may also apply in the internal migration but still remains unknown. Future studies can employ richer population migration data and human capital formation data to empirically check this mechanism.

APPENDICES: SUPPLEMENTARY INFORMATION

Appendix of Chapter 4

A1. The estimation procedures and results of inequality of opportunity

This section presents the detailed calculation procedures of IOP and IOE. First, it is needed to eliminate the effects of age cohorts. This research regresses the natural logarithm of actual income levels on age and age squared following Checchi and Peragine (2010). The regression results are shown in Panel A of Table A1. Then, this research takes the natural exponent of residuals (plus constant) after regression as the outcome for estimating inequality indices. Second, one can calculate inequality indices by the calculated residual income y. Regarding TIE, one can directly calculate the inequality of residual income to represent it by employing the algorithm (G) of mean logarithmic deviation, as shown below:

$$G(y) = \frac{1}{n}\sum_{i=1}^{n} \log \frac{\bar{y}}{y_i}, \tag{A1}$$

where n is the number of observations; $\bar{y}$ is the mean of residual income for all observations. Regarding IOP, RIOP, and IOE, this research first regresses residual income on four circumstance variables (*Gender*, *Maternal schooling years*, *Paternal schooling years*, and *Hukou status*) by each province (the regression results for all observations are shown in Panel B of Table A1). Based on these regression results, one can further obtain the predicted income contributed by circumstances $\tilde{\mu}$. According to Equations 4.3b, 4.3c, and 4.3d, one can calculate the final IOP, RIOP, and IOE. All these calculation procedures can be implemented by the *iop* command in STATA (Juárez & Soloaga, 2014). The calculated inequality indices are presented in the left part (Parametrically) of Table A2.

Table A1. Regression procedures of income for all observations

	All residents	Excluding past migrants
	(1)	(2)
	Panel A: The age effects on ln(income)	
Age	0.096***	0.094***
	(0.007)	(0.007)
Square of age	-0.001***	-0.001***
	(0.000)	(0.000)
Observations	14,062	12,915
Adjusted R^2	0.060	0.057
	Panel B: The effects of circumstances on residual income	
Gender (male=1)	2132.39***	2026.803***
	(122.668)	(121.946)
Maternal schooling years	65.930***	48.098***
	(16.034	(15.940)
Paternal schooling years	84.260***	89.420***
	(19.067)	(19.014)
Hukou status (urban=1)	2920.636***	2745.269***
	(187.048)	(188.195)
Observations	11,712	10,819
Adjusted R^2	0.064	0.061

Notes: 1. the dependent variable in Panel A is the natural logarithm of income, while that in panel B is the residual income; 2. this table only shows the regression results of all observations. In practice, the regression is conducted by each province separately for index calculation; 3. Standard errors are in parentheses; 4. ***, **, and * denote significance at the 1, 5, and 10 percent levels, respectively.

Table A2. Inequality indices across provinces

	Inequality indices (all residents)						
	Parametrically					Non-parametrically	
	Observations	TIE	IOP	RIOP	IOE	IOP	IOE
Beijing	75	0.2976	0.0331	0.1113	0.2645	0.0861	0.2115
Tianjin	113	0.3918	0.0425	0.1085	0.3493	0.0811	0.3107
Hebei	626	0.5971	0.0732	0.1226	0.5239	0.0960	0.5011
Shanxi	507	0.5491	0.0520	0.0947	0.4971	0.0457	0.5034

Liaoning	1002	0.6232	0.0581	0.0932	0.5652	0.0551	0.5681
Jilin	171	0.3522	0.0341	0.0968	0.3181	0.0662	0.2860
Heilongjiang	349	0.3561	0.0146	0.0409	0.3416	0.0297	0.3264
Shanghai	1138	0.3751	0.0438	0.1168	0.3312	0.0474	0.3277
Jiangsu	248	0.3815	0.1042	0.2731	0.2773	0.0810	0.3005
Zhejiang	218	0.3352	0.0571	0.1704	0.2781	0.0457	0.2895
Anhui	182	0.4857	0.0975	0.2008	0.3882	0.0769	0.4088
Fujian	134	0.5379	0.1124	0.2089	0.4255	0.0897	0.4482
Jiangxi	207	0.2806	0.0429	0.1531	0.2376	0.0811	0.1995
Shandong	591	0.4851	0.0590	0.1216	0.4261	0.0555	0.4296
Henan	1206	0.5303	0.0358	0.0675	0.4945	0.0529	0.4774
Hubei	217	0.5474	0.0516	0.0942	0.4958	0.1179	0.4295
Hunan	340	0.6221	0.1607	0.2584	0.4614	0.1312	0.4909
Guangdong	1086	0.5138	0.0552	0.1075	0.4585	0.0680	0.4458
Guangxi	304	0.5570	0.0612	0.1098	0.4958	0.1354	0.4216
Chongqing	89	0.7574	0.4512	0.5957	0.3062	0.3698	0.3876
Sichuan	666	0.6352	0.0961	0.1513	0.5390	0.1735	0.4617
Guizhou	378	0.5962	0.1326	0.2224	0.4636	0.1517	0.4445
Yunnan	351	0.7101	0.0117	0.0164	0.6984	0.0102	0.6999
Shaanxi	292	0.5456	0.0879	0.1611	0.4577	0.2357	0.3099
Gansu	1222	0.7754	0.1749	0.2256	0.6005	0.1064	0.6690
Total	11712	0.6416	0.0837	0.1304	0.5579	0.0841	0.5575

Notes: 1. All inequality indices are calculated by the inequality algorithm (G) of mean logarithmic deviation; 2. because of the data unavailability, provinces do not include Inner Mongolia, Hainan, Tibet, Qinghai, Ningxia, and Xinjiang.

To ensure the robustness of the estimation method utilized in this chapter, this research further tries to use the non-parametric method to calculate inequality indices following Checchi and Peragine (2010) and Marrero and Rodríguez (2013). Given the limitation of samples, it is unable to use four circumstance variables for classification simultaneously. Therefore, only paternal education level and gender are considered as circumstances, similar to Marrero and Rodríguez (2013). First, all observations in each province are divided into eight groups based on four paternal education levels (no education, primary, secondary, and tertiary) and two genders (male and female). Then, the mean residual income in each group is calculated since the ex-ante approach only considers the mean outcomes of groups. Finally, employing the same algorithm shown in Equation A5, the IOP is calculated by the mean logarithmic deviation of eight groups' mean residual incomes (e.g., between-group inequality). In the meantime, the IOE is the difference between TIE (measured in the same way as parametric method) and IOP. All calculated inequality indices for each province are presented in the right part (Non-parametrically) of Table A2. As shown in the table, the indices of the two methods are highly similar, given that the correlation degree of the two IOP indices is 0.834, which is significant at 1 percent level. These results emphasize the stability of inequality indices across various estimation methods.

A2. The specification of the self-selection model

This section will specify the selection model based on utility-maximizing framework. Considering an individual p with skill level s migrate from origin i to destination j, his/her utility is determined by the wage level and the cost of migration. This research assumes a logarithmic-utility model where the utility is the logarithmic function of wage levels $W_{p,i}^{s}$ and migration costs $C_{p,ij}^{s}$ as well as an idiosyncratic term $\varepsilon_{p,ij}^{s}$, such that

$$U_{p,ii}^{s} = \alpha \ln W_{p,i}^{s} + \varepsilon_{p,ij}^{s} \tag{A2a}$$

$$U_{p,ij}^{s} = \alpha \ln\left(W_{p,j}^{s} - C_{p,ij}^{s}\right) + \varepsilon_{p,ij}^{s} = \alpha \ln W_{p,j}^{s} - \pi_{p,ij}^{s} + \varepsilon_{p,ij}^{s}, \tag{A2b}$$

where $\pi_{p,ij}^{s}$ is a "time-equivalent" measure ($\pi_{p,ij}^{s} = C_{p,ij}^{s}/W_{p,j}^{s}$) of migration costs; $\varepsilon_{p,ij}^{s}$ follows an i.i.d. extreme value distribution.

To maximize personal utility, individual p will migrate only if $P_{p,ij}^{s} = Prob\left(U_{p,ij}^{s} > U_{p,ii}^{s}\right) > 0, \forall i \neq j$. Following McFadden (1974), one can employ the log odds of the number of migrants migrating to destination j to the number of stayers at origin i to capture the utility difference between two locations, as shown below:

$$\ln\frac{M_{ij}^{s}}{N_{ii}^{s}} = \alpha\left(\ln W_{j}^{s} - \ln W_{i}^{s}\right) - \alpha\pi_{p,ij}^{s} \tag{A3}$$

where M_{ij}^{s} represents the number of migrants with skill level s from origin i to destination j; N_{ii}^{s} represents the number of natives with skill level s staying at origin i. This equation demonstrates the scale of migration influenced by wage level differentials and migration costs, widely used in the previous literature. This chapter aims to investigate the self-selection of migrants, that is, the migration scale difference between high-skilled and low-skilled individuals. As such, this research divides all population into two skill types (i.e., high-skilled h and low-skilled l), and then make a difference of Equation A2 between these two skill types, as shown below:

$$\ln\frac{M_{ij}^{h}}{N_{ii}^{h}} - \ln\frac{M_{ij}^{l}}{N_{ii}^{l}} = \ln\frac{M_{ij}^{h}/N_{ii}^{h}}{M_{ij}^{l}/N_{ii}^{l}} = \alpha\left[\left(\ln W_{j}^{h} - \ln W_{j}^{l}\right) - \left(\ln W_{i}^{h} - \ln W_{i}^{l}\right)\right]$$

$$-\alpha\left(\pi_{ij}^{h} - \pi_{ij}^{l}\right) \tag{A4}$$

This equation measures the selection of migrants induced by wage differentials between skill types and location in addition to migration costs differentials.

Then, this research further employs the Roy (1951) model to simplify this equation. The basic idea is that personal wage level is determined by the mean wage level μ in location and wage variance ν, such that $\ln W_j^s = \mu_j + \nu_j^s$. Accordingly, one can rewrite the first term in the right-hand of Equation (A3) as $\left[\left(\nu_j^h - \nu_j^l\right) - \left(\nu_i^h - \nu_i^l\right)\right]$. In this equation, $\nu_{j(i)}^h - \nu_{j(i)}^l$ actually reflects the income inequality level $\theta_{j(i)}$ in places of destination and origin. In addition, the second term in the right-hand of Equation (A3) capture migration cost differentials between two skill types, which implies that all non-skill-related migration costs have been offset leaving only skill-related migration costs. Consequently, one can rewrite Equation (A3) as follow:

$$\ln \frac{M_{ij}^h / N_{ii}^h}{M_{ij}^l / N_{ii}^l} = \alpha'\left(\theta_j - \theta_i\right) + \alpha'' \pi'_{ij} \tag{A5}$$

Based on this simplified model, one can derive empirical model as shown in Section 3.4.1. Notably, because the mean wage level and non-skill-related migration costs have been offset, only income inequality and skill-related migration costs can influence migrant selection.

A3. Aggregated emigration rate across provinces

Table A3. Aggregated emigration rate across provinces

	Aggregated emigration rate ($\sum M_{ij}^{H(L)} / N_{ii}^{H(L)}$)			
	All population	High-skilled	Low-skilled	Difference
Beijing	5.01%	4.29%	5.73%	-1.44%
Tianjin	1.03%	1.44%	0.85%	0.59%
Hebei	7.59%	12.48%	7.01%	5.46%
Shanxi	2.61%	3.13%	2.53%	0.60%
Inner Mongolia	1.72%	2.62%	1.54%	1.08%
Liaoning	2.31%	4.71%	1.81%	2.90%
Jilin	2.65%	7.28%	1.96%	5.32%
Heilongjiang	4.43%	10.07%	3.63%	6.43%
Shanghai	1.67%	1.30%	1.92%	-0.62%
Jiangsu	5.78%	7.41%	5.40%	2.02%
Zhejiang	5.36%	5.66%	5.29%	0.37%

Anhui	18.33%	17.78%	18.40%	-0.62%
Fujian	5.14%	4.83%	5.20%	-0.37%
Jiangxi	15.73%	20.35%	15.17%	5.17%
Shandong	4.16%	6.28%	3.86%	2.43%
Henan	6.57%	6.21%	6.61%	-0.40%
Hubei	12.35%	14.73%	11.96%	2.77%
Hunan	16.70%	13.59%	17.19%	-3.60%
Guangdong	1.32%	1.29%	1.33%	-0.04%
Guangxi	13.95%	7.00%	14.73%	-7.73%
Hainan	2.03%	3.19%	1.89%	1.30%
Chongqing	6.93%	4.67%	7.29%	-2.62%
Sichuan	10.46%	7.64%	10.78%	-3.14%
Guizhou	12.70%	4.57%	13.64%	-9.07%
Yunnan	5.85%	2.47%	6.22%	-3.75%
Tibet	0.19%	1.58%	0.13%	1.45%
Shaanxi	4.14%	5.52%	3.92%	1.60%
Gansu	6.42%	7.66%	6.25%	1.41%
Qinghai	0.97%	1.79%	0.86%	0.93%
Ningxia	1.94%	3.24%	1.71%	1.53%
Xinjiang	1.04%	2.16%	0.77%	1.38%
Total	5.97%	5.56%	6.04%	-0.49%

Note: Difference is the difference in emigration rate between high-skilled and low-skilled.

Appendix of Chapter 6

B1. Proof

To compare the migration incentives, we need to compare the wage change of potential migrants after considering housing costs, which can be represented by the difference between A and B:

$$\Delta = B - A = e^{\mu_\pi - \gamma_1 s} - e^{\mu_\pi - \gamma_1 s + \gamma_2 h} = e^{\mu_\pi - \gamma_1 s}(1 - e^{\gamma_2 h}) < 0, \forall\, s \geq 0 \qquad \text{(B1)}$$

such that the migration incentives across all skill groups have been discouraged by the presence of housing costs.

Differentiating equation A1 with respect to s and h, we further have

$$\frac{\partial \Delta}{\partial s} = \gamma_1 e^{\mu_\pi - \gamma_1 s}(e^{\gamma_2 h} - 1) > 0 \qquad \text{(B2)}$$

$$\frac{\partial \Delta}{\partial h} = -\gamma_2 e^{\mu_\pi - \gamma_1 s + \gamma_2 h} < 0. \qquad \text{(B3)}$$

These two equations indicate that potential migrants' wage profile at the destination increases with their skill level and decreases with housing costs at the destination.

Noting that s_L and s_U are two solutions of the following Equation

$$\mu_0 + \eta_0 s = \mu_1 + \eta_1 s(h) - e^{\mu_\pi - \gamma_1 s(h) + \gamma_2 h}. \qquad \text{(B4)}$$

Differentiating equation A4 with respect to h, we have

$$\eta_0 \frac{\partial s}{\partial h} = \eta_1 \frac{\partial s}{\partial h} + \left(\gamma_1 \frac{\partial s}{\partial h} + \gamma_2\right) e^{\mu_\pi - \gamma_1 s + \gamma_2 h}, \tag{B5}$$

and then we get

$$\frac{\partial s}{\partial h} = \frac{\gamma_2 e^{\mu_\pi - \gamma_1 s + \gamma_2 h}}{\eta_0 - \eta_1 - \gamma_1 e^{\mu_\pi - \gamma_1 s + \gamma_2 h}}. \tag{B6}$$

The marginal effect of an increase in housing costs on the two critical skill thresholds is therefore given by

$$\frac{\partial s_L}{\partial h} = \frac{\gamma_2 e^{\mu_\pi - \gamma_1 s_L + \gamma_2 h}}{\eta_0 - \eta_1 - \gamma_1 e^{\mu_\pi - \gamma_1 s_L + \gamma_2 h}} \tag{B7}$$

$$\frac{\partial s_U}{\partial h} = \frac{\gamma_2 e^{\mu_\pi - \gamma_1 s_U + \gamma_2 h}}{\eta_0 - \eta_1 - \gamma_1 e^{\mu_\pi - \gamma_1 s_U + \gamma_2 h}}. \tag{B8}$$

As shown in Figure 6.3, we have $\frac{\partial s_L}{\partial h} > 0$ and $\frac{\partial s_U}{\partial h} < 0$, such that

$$e^{\mu_\pi - \gamma_1 s_L + \gamma_2 h} < (\eta_0 - \eta_1)/\gamma_1 < e^{\mu_\pi - \gamma_1 s_U + \gamma_2 h}$$

To compare values of $|s_L - s_L'|$ and $|s_U - s_U'|$, ONE can directly compare $\partial s_L/\partial h$ and $|\partial s_U/\partial h|$, as shown below

$$\frac{\partial s_L/\partial h}{|\partial s_U/\partial h|} = \frac{\dfrac{\gamma_2 e^{\mu_\pi - \gamma_1 s_L + \gamma_2 h}}{\eta_0 - \eta_1 - \gamma_1 e^{\mu_\pi - \gamma_1 s_L + \gamma_2 h}}}{\left|\dfrac{\gamma_2 e^{\mu_\pi - \gamma_1 s_U + \gamma_2 h}}{\eta_0 - \eta_1 - \gamma_1 e^{\mu_\pi - \gamma_1 s_U + \gamma_2 h}}\right|}$$

$$= \frac{e^{\mu_\pi - \gamma_1 s_L + \gamma_2 h}}{e^{\mu_\pi - \gamma_1 s_U + \gamma_2 h}} \cdot \frac{|\eta_0 - \eta_1 - \gamma_1 e^{\mu_\pi - \gamma_1 s_U + \gamma_2 h}|}{\eta_0 - \eta_1 - \gamma_1 e^{\mu_\pi - \gamma_1 s_L + \gamma_2 h}} > 1,$$

such that

$$|s_L - s_L'| > |s_U - s_U'|.$$

This result proves that increasing housing costs discourage more low-skilled migrants than high-skilled migrants.

B2. Empirical results of different housing indicators

To ensure the robustness of the results, we further discuss other housing cost indicators. Firstly, the housing price level one year before migration is introduced since migrants may make decisions based on historical circumstances. Secondly, some scholars have noticed that the housing price appreciation induces wealth effects on migrants (Peng & Tsai, 2019; Zang et al., 2015). We introduce the housing price growth rate of one and five years into the empirical model. Thirdly, housing affordability is another important indicator of the housing market because it simultaneously considers residents' income levels. Therefore, we use the ratio of housing price to the average wage level of urban workers to represent the housing affordability index. The empirical results are shown in Table 3.

Columns (1) and (5) of Table B1 show the effects of lagged housing prices on migrant education selection, which are still significant and positive. Columns (2), (3), (6), and (7) of Table B1 show the effects of housing price growth on migrant selection. The results demonstrate that whether it is in the short-term or medium- to long-term, housing price growth rates at the destination lead to a significant positive selection of migrants. Housing price appreciation mainly benefits residents that have houses or are able to purchase houses because it increases the value of their wealth. In contrast, housing price growth leads to more tremendous obstacles to low-educated groups' homeownership. Therefore, high housing price growth rate discourages low-educated migrants and induce positive migrant selection. Finally, as shown in columns (4) and (8) of Table B1, we estimate the impact of city-level housing affordability on migrant selection. The differences in the housing affordability index significantly affect migrant selection, which is same as the effect of housing price differences. All the indicators illustrate that higher housing prices and their growth in destination cities induce positive migrant selection and discourage potential low-skilled migrants.

Table B1. Housing prices and migrant selection: other housing indicators

	Years of schooling attended				Obtainment of the college degree (dummy)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Delta ln housing price (t-1)	0.282***				0.172***			
	(0.105)				(0.058)			
Ori housing price growth rate (one year)		-0.122				0.038		
		(0.287)				(0.158)		
Des housing price growth rate (one year)		0.566**				0.525***		
		(0.288)				(0.164)		
Ori housing price growth rate (five years)			-0.039				0.005	
			(0.155)				(0.086)	
Des housing price growth rate (five years)			0.460***				0.363***	
			(0.121)				(0.069)	
Delta housing affordability index				1.082***				0.493***
				(0.115)				(0.054)
Individual-level controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Urban-level controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Origin FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Destination FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Method	OLS	OLS	OLS	OLS	Probit	Probit	Probit	Probit
Observations	13053	13053	13206	15300	13053	13053	13206	15300
Adj R^2	0.299	0.299	0.300	0.313	0.189	0.190	0.191	0.200

Notes: 1. *Delta* means the differences in the destination and original cities; *Ori* and *Des* refer to the variables of the original and destination cities, respectively. 2. ***, **, and * denote significance at the 1, 5, and 10 percent levels, respectively. 3. The standard errors are in parentheses.

REFERENCES

Abramitzky, R., Boustan, L. P., & Eriksson, K. (2012). Europe's Tired, Poor, Huddled Masses: Self-Selection and Economic Outcomes in the Age of Mass Migration. *American Economic Review*, *102*(5), 1832–1856. https://doi.org/10.1257/aer.102.5.1832

Abramitzky, R., Lavy, V., & Pérez, S. (2021). The long-term spillover effects of changes in the return to schooling. *Journal of Public Economics*, *196*, 104369. https://doi.org/10.1016/j.jpubeco.2021.104369

Acemoglu, D., & Restrepo, P. (2018). Modeling Automation. *AEA Papers and Proceedings*, *108*, 48–53. https://doi.org/10.1257/pandp.20181020

Acemoglu, D., & Restrepo, P. (2019). Automation and New Tasks: How Technology Displaces and Reinstates Labor. *Journal of Economic Perspectives*, *33*(2), 3–30. https://doi.org/10.1257/jep.33.2.3

Acemoglu, D., & Restrepo, P. (2020). Robots and Jobs: Evidence from US Labor Markets. *Journal of Political Economy*, *128*(6), 2188–2244. https://doi.org/10.1086/705716

Arrow, K., Bowles, S., & Durlauf, S. N. (2018). *Meritocracy and Economic Inequality*. Princeton University Press.

Asadullah, M. N., & Xiao, S. (2019). Labor Market Returns to Education and English Language Skills in the People's Republic of China: An Update. *Asian Development Review*, *36*(1), 80–111. https://doi.org/10.1162/adev_a_00124

Asadullah, M. N., & Xiao, S. (2020). The changing pattern of wage returns to education in post-reform China. *Structural Change and Economic Dynamics*, *53*, 137–148. https://doi.org/10.1016/j.strueco.2020.01.010

Au, C. C., & Henderson, J. V. (2006). Are Chinese Cities Too Small? *The Review of Economic Studies*, *73*(3), 549–576. https://doi.org/10.1111/j.1467-937X.2006.00387.x

Autor, D. H. (2015). Why Are There Still So Many Jobs? The History and Future of Workplace Automation. *Journal of Economic Perspectives*, *29*(3), 3–30. https://doi.org/10.1257/jep.29.3.3

Autor, D. H., & Dorn, D. (2013). The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market. *American Economic Review*, *103*(5), 1553–1597. https://doi.org/10.1257/aer.103.5.1553

Autor, D. H., Katz, L. F., & Krueger, A. B. (1998). Computing Inequality: Have Computers Changed the Labor Market?*. *The Quarterly Journal of Economics*, *113*(4), 1169–1213. https://doi.org/10.1162/003355398555874

Autor, D. H., Levy, F., & Murnane, R. J. (2003). The Skill Content of Recent Technological

Change: An Empirical Exploration*. *The Quarterly Journal of Economics*, *118*(4), 1279–1333. https://doi.org/10.1162/003355303322552801

Autor, D. H., & Salomons, A. (2018). *Is Automation Labor-Displacing? Productivity Growth, Employment, and the Labor Share* (Working Paper No. 24871; Working Paper Series). National Bureau of Economic Research. https://doi.org/10.3386/w24871

Bacolod, M., De la Roca, J., & Ferreyra, M. M. (2021). In search of better opportunities: Sorting and agglomeration effects among young college graduates in Colombia. *Regional Science and Urban Economics*, *87*, 103656. https://doi.org/10.1016/j.regsciurbeco.2021.103656

Bao, S., Bodvarsson, Ö. B., Hou, J. W., & Zhao, Y. (2011). The Regulation of Migration in a Transition Economy: China's Hukou System. *Contemporary Economic Policy*, *29*(4), 564–579. https://doi.org/10.1111/j.1465-7287.2010.00224.x

Barber, P. G. (2000). Agency in philippine women's labour migration and provisional diaspora. *Women's Studies International Forum*, *23*(4), 399–411. https://doi.org/10.1016/S0277-5395(00)00104-7

Bauernschuster, S., Falck, O., Heblich, S., Suedekum, J., & Lameli, A. (2014). Why are educated and risk-loving persons more mobile across regions? *Journal of Economic Behavior & Organization*, *98*, 56–69. https://doi.org/10.1016/j.jebo.2013.12.011

Behrens, K., Duranton, G., & Robert-Nicoud, F. (2014). Productive Cities: Sorting, Selection, and Agglomeration. *Journal of Political Economy*, *122*(3), 507–553. https://doi.org/10.1086/675534

Beine, M., Docquier, F., & Özden, Ç. (2011). Diasporas. *Journal of Development Economics*, *95*(1), 30–41. https://doi.org/10.1016/j.jdeveco.2009.11.004

Beine, M., Docquier, F., & Rapoport, H. (2001). Brain drain and economic growth: Theory and evidence. *Journal of Development Economics*, *64*(1), 275–289. https://doi.org/10.1016/S0304-3878(00)00133-4

Beine, M., Docquier, F., & Rapoport, H. (2008). Brain Drain and Human Capital Formation in Developing Countries: Winners and Losers. *The Economic Journal*, *118*(528), 631–652. https://doi.org/10.1111/j.1468-0297.2008.02135.x

Belot, M. V. K., & Hatton, T. J. (2012). Immigrant Selection in the OECD*. *The Scandinavian Journal of Economics*, *114*(4), 1105–1128. https://doi.org/10.1111/j.1467-9442.2012.01721.x

Berger, M. C., & Blomquist, G. C. (1992). Mobility and destination in migration decisions: The roles of earnings, quality of life, and housing prices. *Journal of Housing Economics*, *2*(1), 37–59. https://doi.org/10.1016/1051-1377(92)90018-L

Berry, C. R., & Glaeser, E. L. (2005). The divergence of human capital levels across cities*.

Papers in Regional Science, *84*(3), 407–444. https://doi.org/10.1111/j.1435-5957.2005.00047.x

Bertoli, S., Dequiedt, V., & Zenou, Y. (2016). Can selective immigration policies reduce migrants' quality? *Journal of Development Economics*, *119*, 100–109. https://doi.org/10.1016/j.jdeveco.2015.11.002

Björklund, A., Roine, J., & Waldenström, D. (2012). Intergenerational top income mobility in Sweden: Capitalist dynasties in the land of equal opportunity? *Journal of Public Economics*, *96*(5), 474–484. https://doi.org/10.1016/j.jpubeco.2012.02.003

Bloze, G., & Skak, M. (2016). Housing equity, residential mobility and commuting. *Journal of Urban Economics*, *96*, 156–165. https://doi.org/10.1016/j.jue.2016.09.003

Blume, L. E., & Durlauf, S. N. (2000). The interactions-based approach to socioeconomic behavior. In *Working papers* (No. 1; Working Papers). Wisconsin Madison - Social Systems. https://ideas.repec.org/p/att/wimass/20001.html

Bobbitt-Zeher, D. (2007). The Gender Income Gap and the Role of Education. *Sociology of Education*, *80*(1), 1–22. https://doi.org/10.1177/003804070708000101

Borjas, G. J. (1987). Self-Selection and the Earnings of Immigrants. *The American Economic Review*, *77*(4), 531–553.

Borjas, G. J. (1991). Immigration and self-selection. In J. M. Abowd & R. B. Freeman (Eds.), *Immigration, Trade, and the Labor Market* (pp. 29–76). University of Chicago Press.

Borjas, G. J. (2008). Labor Outflows and Labor Inflows in Puerto Rico. *Journal of Human Capital*, *2*(1), 32–68. https://doi.org/10.1086/527521

Borjas, G. J. (2018). The Economics of Immigration. In *Interdisciplinary Perspectives on the New Immigration* (pp. 1–52). Routledge.

Borjas, G. J., & Bratsberg, B. (1996). Who leaves? The outmigration of the foreign-born. *Review of Economics & Statistics*, *78*(1). https://elibrary.ru/item.asp?id=2607564

Borjas, G. J., Bronars, S. G., & Trejo, S. J. (1992). Self-selection and internal migration in the United States. *Journal of Urban Economics*, *32*(2), 159–185. https://doi.org/10.1016/0094-1190(92)90003-4

Borjas, G. J., & Freeman, R. B. (2019). From Immigrants to Robots: The Changing Locus of Substitutes for Workers. *RSF: The Russell Sage Foundation Journal of the Social Sciences*, *5*(5), 22–42. https://doi.org/10.7758/RSF.2019.5.5.02

Borjas, G. J., Kauppinen, I., & Poutvaara, P. (2019). Self-selection of Emigrants: Theory and Evidence on Stochastic Dominance in Observable and Unobservable Characteristics. *The Economic Journal*, *129*(617), 143–171. https://doi.org/10.1111/ecoj.12585

Bourguignon, F., Ferreira, F. H. G., & Menéndez, M. (2007). Inequality of Opportunity in

Brazil. *Review of Income and Wealth*, *53*(4), 585–618. https://doi.org/10.1111/j.1475-4991.2007.00247.x

Bricker, J., & Bucks, B. (2016). Negative home equity, economic insecurity, and household mobility over the Great Recession. *Journal of Urban Economics*, *91*, 1–12. https://doi.org/10.1016/j.jue.2015.10.001

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. https://doi.org/10.1016/j.regsciurbeco.2015.08.003

Cai, F., Du, Y., & Wang, M. (2001). Household registration system and labor market protection. *Economic Research Journal*, *12*, 41–49.

Cao, Z., Zheng, X., Liu, Y., Li, Y., & Chen, Y. (2018). Exploring the changing patterns of China's migration and its determinants using census data of 2000 and 2010. *Habitat International*, *82*, 72–82. https://doi.org/10.1016/j.habitatint.2018.09.006

Chan, K. W. (2009). The Chinese Hukou System at 50. *Eurasian Geography and Economics*, *50*(2), 197–221. https://doi.org/10.2747/1539-7216.50.2.197

Chan, K. W. (2012). Migration and development in China: Trends, geography and current issues. *Migration and Development*, *1*(2), 187–205. https://doi.org/10.1080/21632324.2012.739316

Chan, K. W., & Buckingham, W. (2008). Is China Abolishing the Hukou System?*. *The China Quarterly*, *195*, 582–606. https://doi.org/10.1017/S0305741008000787

Chan, K. W., & Zhang, L. (1999). The Hukou System and Rural-Urban Migration in China: Processes and Changes*. *The China Quarterly*, *160*, 818–855. https://doi.org/10.1017/S0305741000001351

Chan, S. (2001). Spatial Lock-in: Do Falling House Prices Constrain Residential Mobility? *Journal of Urban Economics*, *49*(3), 567–586. https://doi.org/10.1006/juec.2000.2205

Chang, H., Dong, X., & 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. https://doi.org/10.1016/j.worlddev.2011.05.021

Checchi, D., & Peragine, V. (2010). Inequality of opportunity in Italy. *The Journal of Economic Inequality*, *8*(4), 429–450. https://doi.org/10.1007/s10888-009-9118-3

Chen, B., Liu, D., & Lu, M. (2018). City size, migration and urban inequality in China. *China Economic Review*, *51*, 42–58. https://doi.org/10.1016/j.chieco.2018.05.001

Chen, J., Hao, Q., & Stephens, M. (2010). Assessing Housing Affordability in Post-reform China: A Case Study of Shanghai. *Housing Studies*, *25*(6), 877–901. https://doi.org/10.1080/02673037.2010.511153

Chen, J., Hu, M., & Lin, Z. (2019). Does housing unaffordability crowd out elites in Chinese superstar cities? *Journal of Housing Economics*, *45*, 101571. https://doi.org/10.1016/j.jhe.2018.03.003

Chen, S., Oliva, P., & Zhang, P. (2022). The effect of air pollution on migration: Evidence from China. *Journal of Development Economics*, *156*, 102833. https://doi.org/10.1016/j.jdeveco.2022.102833

Cheng, H., Jia, R., Li, D., & Li, H. (2019). The Rise of Robots in China. *Journal of Economic Perspectives*, *33*(2), 71–88. https://doi.org/10.1257/jep.33.2.71

Chiquiar, D., & Hanson, G. H. (2005). International Migration, Self-Selection, and the Distribution of Wages: Evidence from Mexico and the United States. *Journal of Political Economy*, *113*(2), 239–281. https://doi.org/10.1086/427464

Chiu, W. H. (1998). Income Inequality, Human Capital Accumulation and Economic Performance. *The Economic Journal*, *108*(446), 44–59. https://doi.org/10.1111/1468-0297.00272

Clark, X., Hatton, T. J., & Williamson, J. G. (2007). Explaining U.S. Immigration, 1971–1998. *The Review of Economics and Statistics*, *89*(2), 359–373. https://doi.org/10.1162/rest.89.2.359

Combes, P.-P., Duranton, G., & Gobillon, L. (2008). Spatial wage disparities: Sorting matters! *Journal of Urban Economics*, *63*(2), 723–742. https://doi.org/10.1016/j.jue.2007.04.004

Combes, P.-P., Duranton, G., Gobillon, L., & Roux, S. (2012). Sorting and local wage and skill distributions in France. *Regional Science and Urban Economics*, *42*(6), 913–930. https://doi.org/10.1016/j.regsciurbeco.2012.11.003

Crowley, F., Doran, J., & McCann, P. (2021). The vulnerability of European regional labour markets to job automation: The role of agglomeration externalities. *Regional Studies*, *55*(10–11), 1711–1723. https://doi.org/10.1080/00343404.2021.1928041

Czaika, M., & Parsons, C. R. (2017). The Gravity of High-Skilled Migration Policies. *Demography*, *54*(2), 603–630. https://doi.org/10.1007/s13524-017-0559-1

Dardanoni, V., Fields, G. S., Roemer, J. E., & Puerta, M. L. S. (2006). How Demanding Should Equality of Opportunity Be, and How Much. In S. L. Morgan, D. B. Grusky, & G. S. Fields (Eds.), *Mobility and Inequality: Frontiers of Research in Sociology and Economics*. Stanford University Press.

Dauth, W., Findeisen, S., Suedekum, J., & Woessner, N. (2021). The Adjustment of Labor Markets to Robots. *Journal of the European Economic Association*, *19*(6), 3104–3153. https://doi.org/10.1093/jeea/jvab012

Davies, P. S., Greenwood, M. J., & Li, H. (2001). A Conditional Logit Approach to U.S. State-to-State Migration. *Journal of Regional Science*, *41*(2), 337–360.

https://doi.org/10.1111/0022-4146.00220

de Haas, H., Natter, K., & Vezzoli, S. (2016). Growing Restrictiveness or Changing Selection? The Nature and Evolution of Migration Policies. *International Migration Review*, 1–44. https://doi.org/10.1111/imre.12288

De la Roca, J. (2017). Selection in initial and return migration: Evidence from moves across Spanish cities. *Journal of Urban Economics*, *100*, 33–53. https://doi.org/10.1016/j.jue.2017.04.004

de Lange, T., Berntsen, L., Hanoeman, R., & Haidar, O. (2021). Highly Skilled Entrepreneurial Refugees: Legal and Practical Barriers and Enablers to Start Up in the Netherlands. *International Migration*, *59*(4), 74–87. https://doi.org/10.1111/imig.12745

de Vries, G. J., Gentile, E., Miroudot, S., & Wacker, K. M. (2020). The rise of robots and the fall of routine jobs. *Labour Economics*, *66*, 101885. https://doi.org/10.1016/j.labeco.2020.101885

Diamond, R. (2016). The Determinants and Welfare Implications of US Workers' Diverging Location Choices by Skill: 1980-2000. *American Economic Review*, *106*(3), 479–524. https://doi.org/10.1257/aer.20131706

Du, Y., Park, A., & Wang, S. (2005). Migration and rural poverty in China. *Journal of Comparative Economics*, *33*(4), 688–709. https://doi.org/10.1016/j.jce.2005.09.001

Du, Y., & Wei, X. (2021). Technological change and unemployment: Evidence from China. *Applied Economics Letters*, *0*(0), 1–4. https://doi.org/10.1080/13504851.2021.1896666

Durand, J., & Massey, D. S. (1992). Mexican Migration to the United States: A Critical Review. *Latin American Research Review*, *27*(2), 3–42.

Duranton, G., & Puga, D. (2004). Micro-Foundations of Urban Agglomeration Economies. In J. V. Henderson & J.-F. Thisse (Eds.), *Handbook of Regional and Urban Economics* (Vol. 4, pp. 2063–2117). Elsevier. https://doi.org/10.1016/S1574-0080(04)80005-1

Eeckhout, J., Pinheiro, R., & Schmidheiny, K. (2014). Spatial Sorting. *Journal of Political Economy*, *122*(3), 554–620. https://doi.org/10.1086/676141

Elvery, J. A. (2010). City size and skill intensity. *Regional Science and Urban Economics*, *40*(6), 367–379. https://doi.org/10.1016/j.regsciurbeco.2010.05.006

Fan, C. C. (2005). Modeling Interprovincial Migration in China, 1985-2000. *Eurasian Geography and Economics*, *46*(3), 165–184. https://doi.org/10.2747/1538-7216.46.3.165

Fan, Y., Yi, J., & Zhang, J. (2021). Rising Intergenerational Income Persistence in China. *American Economic Journal: Economic Policy*, *13*(1), 202–230. https://doi.org/10.1257/pol.20170097

Ferreira, F. H. G., & Gignoux, J. (2011). The Measurement of Inequality of Opportunity:

Theory and an Application to Latin America. *Review of Income and Wealth*, *57*(4), 622–657. https://doi.org/10.1111/j.1475-4991.2011.00467.x

Foote, A. (2016). The effects of negative house price changes on migration: Evidence across U.S. housing downturns. *Regional Science and Urban Economics*, *60*, 292–299. https://doi.org/10.1016/j.regsciurbeco.2016.08.001

Fratesi, U., & Percoco, M. (2014). Selective Migration, Regional Growth and Convergence: Evidence from Italy. *Regional Studies*, *48*(10), 1650–1668. https://doi.org/10.1080/00343404.2013.843162

Freeman, C. P. (2015). *Handbook on China and Developing Countries*. Edward Elgar Publishing.

Frey, C. B., & Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change*, *114*, 254–280. https://doi.org/10.1016/j.techfore.2016.08.019

Fu, Q., & Ren, Q. (2010). Educational Inequality under China's Rural–Urban Divide: The Hukou System and Return to Education. *Environment and Planning A: Economy and Space*, *42*(3), 592–610. https://doi.org/10.1068/a42101

Gabriel, S. A., Shack-Marquez, J., & Wascher, W. L. (1992). Regional house-price dispersion and interregional migration. *Journal of Housing Economics*, *2*(3), 235–256. https://doi.org/10.1016/1051-1377(92)90002-8

Gallie, D. (1991). Patterns of Skill Change: Upskilling, Deskilling or the Polarization of Skills? *Work, Employment and Society*, *5*(3), 319–351. https://doi.org/10.1177/0950017091005003002

Galor, O., & Zeira, J. (1993). Income Distribution and Macroeconomics. *The Review of Economic Studies*, *60*(1), 35–52. https://doi.org/10.2307/2297811

Gamboa, L. F., & Waltenberg, F. D. (2012). Inequality of opportunity for educational achievement in Latin America: Evidence from PISA 2006–2009. *Economics of Education Review*, *31*(5), 694–708. https://doi.org/10.1016/j.econedurev.2012.05.002

Gang, I. N., & Rivera-Batiz, F. L. (1994). Labor market effects of immigration in the United States and Europe. *Journal of Population Economics*, *7*(2), 157–175. https://doi.org/10.1007/BF00173617

Ganong, P., & Shoag, D. (2017). Why has regional income convergence in the U.S. declined? *Journal of Urban Economics*, *102*, 76–90. https://doi.org/10.1016/j.jue.2017.07.002

Giuntella, O., & Wang, T. (2019). *Is an Army of Robots Marching on Chinese Jobs?* (SSRN Scholarly Paper ID 3390271). Social Science Research Network. https://doi.org/10.2139/ssrn.3390271

Glaeser, E. L., & Maré, D. C. (2001). Cities and Skills. *Journal of Labor Economics*, *19*(2), 316–342. https://doi.org/10.1086/319563

Glaeser, E. L., & Resseger, M. G. (2010). The Complementarity Between Cities and Skills*. *Journal of Regional Science*, *50*(1), 221–244. https://doi.org/10.1111/j.1467-9787.2009.00635.x

Golley, J., & Kong, S. T. (2013). Inequality in Intergenerational Mobility of Education in China. *China & World Economy*, *21*(2), 15–37. https://doi.org/10.1111/j.1749-124X.2013.12013.x

Goos, M., & Manning, A. (2007). Lousy and Lovely Jobs: The Rising Polarization of Work in Britain. *The Review of Economics and Statistics*, *89*(1), 118–133. https://doi.org/10.1162/rest.89.1.118

Gould, E. D., & Moav, O. (2016). Does High Inequality Attract High Skilled Immigrants? *The Economic Journal*, *126*(593), 1055–1091. https://doi.org/10.1111/ecoj.12185

Graetz, G., & Michaels, G. (2018). Robots at Work. *The Review of Economics and Statistics*, *100*(5), 753–768. https://doi.org/10.1162/rest_a_00754

Gravemeyer, S., Gries, T., & Xue, J. (2011). Income Determination and Income Discrimination in Shenzhen. *Urban Studies*, *48*(7), 1457–1475. https://doi.org/10.1177/0042098010375317

Grogger, J., & Hanson, G. H. (2011). Income maximization and the selection and sorting of international migrants. *Journal of Development Economics*, *95*(1), 42–57. https://doi.org/10.1016/j.jdeveco.2010.06.003

Gyourko, J., Mayer, C., & Sinai, T. (2013). Superstar Cities. *American Economic Journal: Economic Policy*, *5*(4), 167–199. https://doi.org/10.1257/pol.5.4.167

Haddad, M. (2020). When states encourage migration. The institutionalisation of French overseas-mainland migration and its effect on migrant selection. *Journal of Ethnic and Migration Studies*, *0*(0), 1–19. https://doi.org/10.1080/1369183X.2020.1782179

Han, J., Liu, R., & Zhang, J. (2012). Globalization and wage inequality: Evidence from urban China. *Journal of International Economics*, *87*(2), 288–297. https://doi.org/10.1016/j.jinteco.2011.12.006

Hao, T., Sun, R., Tombe, T., & Zhu, X. (2020). The effect of migration policy on growth, structural change, and regional inequality in China. *Journal of Monetary Economics*, *113*, 112–134. https://doi.org/10.1016/j.jmoneco.2020.03.003

Harris, J. R., & Todaro, M. P. (1970). Migration, Unemployment and Development: A Two-Sector Analysis. *The American Economic Review*, *60*(1), 126–142.

Hendricks, L. (2011). The Skill Composition of U.s. Cities*. *International Economic Review*, *52*(1), 1–32. https://doi.org/10.1111/j.1468-2354.2010.00617.x

Huang, H., & Tang, Y. (2012). Residential land use regulation and the US housing price cycle between 2000 and 2009. *Journal of Urban Economics*, *71*(1), 93–99.

https://doi.org/10.1016/j.jue.2011.08.001

Huang, Y., Guo, F., & Tang, Y. (2010). Hukou status and social exclusion of rural–urban migrants in transitional China. *Journal of Asian Public Policy*, *3*(2), 172–185. https://doi.org/10.1080/17516234.2010.501160

Hui, E. C. M., & Gu, Q. (2009). Study of guangzhou house price bubble based on state-space model. *International Journal of Strategic Property Management*, *13*(4), 287–298. https://doi.org/10.3846/1648-715X.2009.13.287-298

Hui, E. C. M., Yu, K. H., & Ye, Y. (2014). Housing Preferences of Temporary Migrants in Urban China in the wake of Gradual Hukou Reform: A Case Study of Shenzhen. *International Journal of Urban and Regional Research*, *38*(4), 1384–1398. https://doi.org/10.1111/j.1468-2427.2012.01176.x

Jacobs, J. A. (1996). Gender Inequality and Higher Education. *Annual Review of Sociology*, *22*(1), 153–185. https://doi.org/10.1146/annurev.soc.22.1.153

Jaschke, P., & Keita, S. (2021). Say it like Goethe: Language learning facilities abroad and the self-selection of immigrants. *Journal of Development Economics*, *149*, 102597. https://doi.org/10.1016/j.jdeveco.2020.102597

Jeanty, P. W., Partridge, M., & Irwin, E. (2010). Estimation of a spatial simultaneous equation model of population migration and housing price dynamics. *Regional Science and Urban Economics*, *40*(5), 343–352. https://doi.org/10.1016/j.regsciurbeco.2010.01.002

Juárez, F. W. C., & Soloaga, I. (2014). Iop: Estimating Ex-Ante Inequality of Opportunity. *The Stata Journal*, *14*(4), 830–846. https://doi.org/10.1177/1536867X1401400408

Kan, K. (2003). Residential mobility and job changes under uncertainty. *Journal of Urban Economics*, *54*(3), 566–586. https://doi.org/10.1016/S0094-1190(03)00086-X

Kanbur, R., & Rapoport, H. (2005). Migration selectivity and the evolution of spatial inequality. *Journal of Economic Geography*, *5*(1), 43–57. https://doi.org/10.1093/jnlecg/lbh053

Katz, L. F., & Autor, D. H. (1999). Chapter 26—Changes in the Wage Structure and Earnings Inequality. In O. C. Ashenfelter & D. Card (Eds.), *Handbook of Labor Economics* (Vol. 3, pp. 1463–1555). Elsevier. https://doi.org/10.1016/S1573-4463(99)03007-2

Katz, L. F., & Murphy, K. M. (1992). Changes in Relative Wages, 1963–1987: Supply and Demand Factors*. *The Quarterly Journal of Economics*, *107*(1), 35–78. https://doi.org/10.2307/2118323

Kinnan, C., Wang, S.-Y., & Wang, Y. (2018). Access to Migration for Rural Households. *American Economic Journal: Applied Economics*, *10*(4), 79–119. https://doi.org/10.1257/app.20160395

Koslowski, R. (2014). Selective Migration Policy Models and Changing Realities of Implementation. *International Migration*, *52*(3), 26–39. https://doi.org/10.1111/imig.12136

Lang, K., & Kahn-Lang Spitzer, A. (2020). Race Discrimination: An Economic Perspective. *Journal of Economic Perspectives*, *34*(2), 68–89. https://doi.org/10.1257/jep.34.2.68

Lefranc, A., Pistolesi, N., & Trannoy, A. (2008). Inequality of Opportunities Vs. Inequality of Outcomes: Are Western Societies All Alike? *Review of Income and Wealth*, *54*(4), 513–546. https://doi.org/10.1111/j.1475-4991.2008.00289.x

Leigh, N. G., & Kraft, B. R. (2018). Emerging robotic regions in the United States: Insights for regional economic evolution. *Regional Studies*, *52*(6), 804–815. https://doi.org/10.1080/00343404.2016.1269158

Lewis, W. A. (1954). Economic Development with Unlimited Supplies of Labour. *The Manchester School*, *22*(2), 139–191. https://doi.org/10.1111/j.1467-9957.1954.tb00021.x

Li, B. (2018). Export expansion, skill acquisition and industry specialization: Evidence from china. *Journal of International Economics*, *114*, 346–361. https://doi.org/10.1016/j.jinteco.2018.07.009

Li, J., & Liu, Z. (2018). Housing stress and mental health of migrant populations in urban China. *Cities*, *81*, 172–179. https://doi.org/10.1016/j.cities.2018.04.006

Li, K., Qin, Y., & Wu, J. (2020). Recent housing affordability in urban China: A comprehensive overview. *China Economic Review*, *59*, 101362. https://doi.org/10.1016/j.chieco.2019.101362

Li, X., Zheng, S., Qing, X., Qing, Y., Xu, W., Huang, Y., & Huang, M. (2019). *Recreating New Advantages-How does the urbanization of the Pearl River Delta respond to the "machine generation."* China Regional Coordinated Development and Rural Construction Institute, Sun Yat-sen University.

Liang, W., Lu, M., & Zhang, H. (2016). Housing prices raise wages: Estimating the unexpected effects of land supply regulation in China. *Journal of Housing Economics*, *33*, 70–81. https://doi.org/10.1016/j.jhe.2016.07.002

Liang, Z. (2001). The Age of Migration in China. *Population and Development Review*, *27*(3), 499–524. https://doi.org/10.1111/j.1728-4457.2001.00499.x

Lin, X., Ren, T., Wu, H., & Xiao, Y. (2021). Housing price, talent movement, and innovation output: Evidence from Chinese cities. *Review of Development Economics*, *25*(1), 76–103. https://doi.org/10.1111/rode.12705

Liu, Y., & Shen, J. (2014). Spatial patterns and determinants of skilled internal migration in China, 2000–2005. *Papers in Regional Science*, *93*(4), 749–771. https://doi.org/10.1111/pirs.12014

Liu, Z., Wang, Y., & Chen, S. (2017). Does formal housing encourage settlement intention of rural migrants in Chinese cities? A structural equation model analysis. *Urban Studies*, *54*(8), 1834–1850. https://doi.org/10.1177/0042098016634979

Loury, G. (1989). Why should we care about inequality? In S. Shulman & W. Darity (Eds.), *The Question of Discrimination: Racial Inequality in the U.S. Labor Market* (pp. 268–290). Wesleyan University Press.

Lu, Y. (2012). Education of Children Left Behind in Rural China. *Journal of Marriage and Family*, *74*(2), 328–341. https://doi.org/10.1111/j.1741-3737.2011.00951.x

Lucas, R. E. (1988). On the mechanics of economic development. *Journal of Monetary Economics*, *22*(1), 3–42. https://doi.org/10.1016/0304-3932(88)90168-7

Luo, C., Li, S., & Sicular, T. (2020). The long-term evolution of national income inequality and rural poverty in China. *China Economic Review*, *62*, 101465. https://doi.org/10.1016/j.chieco.2020.101465

Luo, D., & Xing, C. (2016). Population adjustments in response to local demand shifts in China. *Journal of Housing Economics*, *33*, 101–114. https://doi.org/10.1016/j.jhe.2016.05.001

Lux, M., & Sunega, P. (2012). Labour Mobility and Housing: The Impact of Housing Tenure and Housing Affordability on Labour Migration in the Czech Republic. *Urban Studies*, *49*(3), 489–504. https://doi.org/10.1177/0042098011405693

Marrero, G. A., & Rodríguez, J. G. (2013). Inequality of opportunity and growth. *Journal of Development Economics*, *104*, 107–122. https://doi.org/10.1016/j.jdeveco.2013.05.004

Martin, S. (2007). *Women, migration and development* (Policy Brief 1; Transatlantic Perspectives on Migration). Institute for the Study of International Migration.

Mayda, A. M. (2010). International migration: A panel data analysis of the determinants of bilateral flows. *Journal of Population Economics*, *23*(4), 1249–1274. https://doi.org/10.1007/s00148-009-0251-x

McFadden, D. (1974). The measurement of urban travel demand. *Journal of Public Economics*, *3*(4), 303–328. https://doi.org/10.1016/0047-2727(74)90003-6

McKenzie, D., & Rapoport, H. (2010). Self-Selection Patterns in Mexico-U.S. Migration: The Role of Migration Networks. *The Review of Economics and Statistics*, *92*(4), 811–821. https://doi.org/10.1162/REST_a_00032

Mejía, D., & St-Pierre, M. (2008). Unequal opportunities and human capital formation. *Journal of Development Economics*, *86*(2), 395–413. https://doi.org/10.1016/j.jdeveco.2007.04.001

Meng, X. (2012). Labor Market Outcomes and Reforms in China. *Journal of Economic Perspectives*, *26*(4), 75–102. https://doi.org/10.1257/jep.26.4.75

Meng, X., & Manning, C. (2010). The Great Migration in China and Indonesia: Trends and Institutions. In X. Meng, C. Manning, L. Shi, & T. N. Effendi (Eds.), *The Great Migration in China and Indonesia* (Vol. 1, pp. 1–22). Edward Elgar Publishing. https://www.elgaronline.com/view/edcoll/9781848446441/9781848446441.00008.xml

Michaelides, M. (2011). The effect of local ties, wages, and housing costs on migration decisions. *The Journal of Socio-Economics*, *40*(2), 132–140. https://doi.org/10.1016/j.socec.2011.01.010

Mincer, J. (1970). The Distribution of Labor Incomes: A Survey With Special Reference to the Human Capital Approach. *Journal of Economic Literature*, *8*(1), 1–26.

Mincer, J. (1974). *Schooling, Experience, and Earnings*. Human Behavior & Social Institutions No. 2.

Molloy, R., Smith, C. L., & Wozniak, A. (2011). Internal Migration in the United States. *Journal of Economic Perspectives*, *25*(3), 173–196. https://doi.org/10.1257/jep.25.3.173

Murphy, K. M., Riddell, W. C., & Romer, P. M. (1998). *Wages, Skills, and Technology in the United States and Canada* (Working Paper No. 6638; Working Paper Series). National Bureau of Economic Research. https://doi.org/10.3386/w6638

Østbye, S., & Westerlund, O. (2007). Is Migration Important for Regional Convergence? Comparative Evidence for Norwegian and Swedish Counties, 1980–2000. *Regional Studies*, *41*(7), 901–915. https://doi.org/10.1080/00343400601142761

Palomino, J. C., Marrero, G. A., & Rodríguez, J. G. (2019). Channels of Inequality of Opportunity: The Role of Education and Occupation in Europe. *Social Indicators Research*, *143*(3), 1045–1074. https://doi.org/10.1007/s11205-018-2008-y

Peng, C.-W., & Tsai, I.-C. (2019). The long- and short-run influences of housing prices on migration. *Cities*, *93*, 253–262. https://doi.org/10.1016/j.cities.2019.05.011

Plane, D. A. (1993). Demographic Influences on Migration. *Regional Studies*, *27*(4), 375–383. https://doi.org/10.1080/00343409312331347635

Potepan, M. J. (1994). Intermetropolitan Migration and Housing Prices: Simultaneously Determined? *Journal of Housing Economics*, *3*(2), 77–91. https://doi.org/10.1006/jhec.1994.1001

Puga, D. (2010). The Magnitude and Causes of Agglomeration Economies*. *Journal of Regional Science*, *50*(1), 203–219. https://doi.org/10.1111/j.1467-9787.2009.00657.x

Qian, H., & Walker, A. (2015). The education of migrant children in Shanghai: The battle for equity. *International Journal of Educational Development*, *44*, 74–81. https://doi.org/10.1016/j.ijedudev.2015.07.009

Rabe, B., & Taylor, M. P. (2012). Differences in Opportunities? Wage, Employment and House-Price Effects on Migration*. *Oxford Bulletin of Economics and Statistics*, *74*(6), 831–855. https://doi.org/10.1111/j.1468-0084.2011.00682.x

Rawls, J. (1971). A Theory of Justice. In *A Theory of Justice*. Harvard University Press. https://doi.org/10.4159/9780674042605

Roback, J. (1982). Wages, Rents, and the Quality of Life. *Journal of Political Economy*, *90*(6), 1257–1278. https://doi.org/10.1086/261120

Roca, J. D. L., & Puga, D. (2017). Learning by Working in Big Cities. *The Review of Economic Studies*, *84*(1), 106–142. https://doi.org/10.1093/restud/rdw031

Roemer, J. E. (1998). Equality of Opportunity. In *Equality of Opportunity*. Harvard University Press. https://doi.org/10.4159/9780674042872

Rosen, S. (1979). Wage-Based Indexes of Urban Quality of Life. *Current Issues in Urban Economics*, 74–104.

Roy, A. D. (1951). Some Thoughts on the Distribution of Earnings on JSTOR. *Oxford Economic Papers*, *3*(2), 135–146.

Sen, A. (1980). Equality of What? In *The Tanner Lecture on Human Values: Vol. I* (pp. 197–220). Cambridge University Press. https://ictlogy.net/bibliography/reports/projects.php?idp=845&lang=en

Shen, J., & Liu, Y. (2016). Skilled and less-skilled interregional migration in China: A comparative analysis of spatial patterns and the decision to migrate in 2000–2005. *Habitat International*, *57*, 1–10. https://doi.org/10.1016/j.habitatint.2016.06.007

Shi, W., Chen, J., & Wang, H. (2016). Affordable housing policy in China: New developments and new challenges. *Habitat International*, *54*, 224–233. https://doi.org/10.1016/j.habitatint.2015.11.020

Sicular, T., Yue, X., Bjorn, G., & Li, S. (2007). The urban-rural income gap and inequality in China. *Review of Income and Wealth*, *53*(1), 93–126.

Sieg, H., Yoon, C., & Zhang, J. (2020). *The Impact of Migration Controls on Urban Fiscal Policies and the Intergenerational Transmission of Human Capital in China* (Working Paper No. 27764; Working Paper Series). National Bureau of Economic Research. https://doi.org/10.3386/w27764

Sjaastad, L. A. (1962). The Costs and Returns of Human Migration. *Journal of Political Economy*, *70*(5, Part 2), 80–93. https://doi.org/10.1086/258726

Song, Q., & Smith, J. P. (2021). The Citizenship Advantage in Psychological Well-being: An Examination of the Hukou System in China. *Demography*, *58*(1), 165–189. https://doi.org/10.1215/00703370-8913024

Song, Y. (2014). What should economists know about the current Chinese hukou system? *China Economic Review*, *29*, 200–212. https://doi.org/10.1016/j.chieco.2014.04.012

Song, Y. (2016). Hukou-based labour market discrimination and ownership structure in urban China. *Urban Studies*, *53*(8), 1657–1673. https://doi.org/10.1177/0042098015576861

Song, Y., Wu, W., & Zhou, G. (2020). Inequality of opportunity and household risky asset investment: Evidence from panel data in China. *China Economic Review*, *63*, 101513. https://doi.org/10.1016/j.chieco.2020.101513

Song, Y., & Zhang, C. (2020). City size and housing purchase intention: Evidence from rural–urban migrants in China. *Urban Studies*, *57*(9), 1866–1886. https://doi.org/10.1177/0042098019856822

Song, Y., & Zhou, G. (2019). Inequality of opportunity and household education expenditures: Evidence from panel data in China. *China Economic Review*, *55*, 85–98. https://doi.org/10.1016/j.chieco.2019.03.002

Storper, M. (2018). Separate Worlds? Explaining the current wave of regional economic polarization. *Journal of Economic Geography*, *18*(2), 247–270. https://doi.org/10.1093/jeg/lby011

Tao, L., Hui, E. C. M., Wong, F. K. W., & Chen, T. (2015). Housing choices of migrant workers in China: Beyond the Hukou perspective. *Habitat International*, *49*, 474–483. https://doi.org/10.1016/j.habitatint.2015.06.018

Tao, L., Wong, F. K. W., & Hui, E. C. M. (2014). Residential satisfaction of migrant workers in China: A case study of Shenzhen. *Habitat International*, *42*, 193–202. https://doi.org/10.1016/j.habitatint.2013.12.006

Todaro, M. P. (1969). A Model of Labor Migration and Urban Unemployment in Less Developed Countries on JSTOR. *The American Economic Review*, *59*(1), 139–148.

Torche, F. (2015). Intergenerational Mobility and Equality of Opportunity. *European Journal of Sociology / Archives Européennes de Sociologie*, *56*(3), 343–371. https://doi.org/10.1017/S0003975615000181

Van De Gaer, D. (1993). *Equality of opportunity and investment in human capital.* [Leuven University]. https://elibrary.ru/item.asp?id=6863044

Van Oort, F. G. (2007). Spatial and sectoral composition effects of agglomeration economies in the Netherlands*. *Papers in Regional Science*, *86*(1), 5–30. https://doi.org/10.1111/j.1435-5957.2006.00088.x

Wang, S., & Hu, Y. (2019). Migration and health in China: Linking sending and host societies. *Population, Space and Place*, *25*(6), e2231. https://doi.org/10.1002/psp.2231

Wang, W. W., & Fan, C. C. (2006). Success or Failure: Selectivity and Reasons of Return Migration in Sichuan and Anhui, China. *Environment and Planning A: Economy and Space*, *38*(5), 939–958. https://doi.org/10.1068/a37428

Wang, W. W., & Fan, C. C. (2012). Migrant Workers' Integration in Urban China: Experiences in Employment, Social Adaptation, and Self-Identity. *Eurasian Geography and Economics*, *53*(6), 731–749. https://doi.org/10.2747/1539-7216.53.6.731

Wang, X.-R., Hui, E. C.-M., Choguill, C., & Jia, S.-H. (2015). The new urbanization policy in

China: Which way forward? *Habitat International*, *47*, 279–284. https://doi.org/10.1016/j.habitatint.2015.02.001

Wang, X.-R., Hui, E. C.-M., & Sun, J.-X. (2017). Population migration, urbanization and housing prices: Evidence from the cities in China. *Habitat International*, *66*, 49–56. https://doi.org/10.1016/j.habitatint.2017.05.010

Ward, Z. (2020). Internal Migration, Education, and Intergenerational Mobility: Evidence from American History. *Journal of Human Resources*, 0619. https://doi.org/10.3368/jhr.58.2.0619-10265R2

Whalley, J., & Xing, C. (2014). The regional distribution of skill premia in urban China: Implications for growth and inequality. *International Labour Review*, *153*(3), 395–419. https://doi.org/10.1111/j.1564-913X.2014.00009.x

Wu, J., Yu, Z., Wei, Y. D., & Yang, L. (2019). Changing distribution of migrant population and its influencing factors in urban China: Economic transition, public policy, and amenities. *Habitat International*, *94*, 102063. https://doi.org/10.1016/j.habitatint.2019.102063

Wu, Z. (2010). Self-selection and Earnings of Migrants: Evidence from Rural China. *Asian Economic Journal*, *24*(1), 23–44. https://doi.org/10.1111/j.1467-8381.2010.02028.x

Xie, S., & Chen, J. (2018). Beyond homeownership: Housing conditions, housing support and rural migrant urban settlement intentions in China. *Cities*, *78*, 76–86. https://doi.org/10.1016/j.cities.2018.01.020

Xie, Y., & Hu, J. (2014). An Introduction to the China Family Panel Studies (CFPS). *Chinese Sociological Review*, *47*(1), 3–29. https://doi.org/10.2753/CSA2162-0555470101.2014.11082908

Xing, C. (2014). Migration, self-selection and income distributions. *Economics of Transition and Institutional Change*, *22*(3), 539–576. https://doi.org/10.1111/ecot.12041

Yang, S., & Guo, F. (2018). Breaking the barriers: How urban housing ownership has changed migrants' settlement intentions in China. *Urban Studies*, *55*(16), 3689–3707. https://doi.org/10.1177/0042098018757873

Yang, Z., & Pan, Y. (2020). Human capital, housing prices, and regional economic development: Will "vying for talent" through policy succeed? *Cities*, *98*, 102577. https://doi.org/10.1016/j.cities.2019.102577

Ye, J., & Lu, P. (2011). Differentiated childhoods: Impacts of rural labor migration on left-behind children in China. *The Journal of Peasant Studies*, *38*(2), 355–377. https://doi.org/10.1080/03066150.2011.559012

Zang, B., Lv, P., & Warren, C. M. J. (2015). Housing prices, rural–urban migrants' settlement decisions and their regional differences in China. *Habitat International*, *50*, 149–159. https://doi.org/10.1016/j.habitatint.2015.08.003

Zeng, H., Yu, X., & Zhang, J. (2019). Urban village demolition, migrant workers' rental costs and housing choices: Evidence from Hangzhou, China. *Cities*, *94*, 70–79. https://doi.org/10.1016/j.cities.2019.05.029

Zhang, C., Jia, S., & Yang, R. (2016). Housing affordability and housing vacancy in China: The role of income inequality. *Journal of Housing Economics*, *33*, 4–14. https://doi.org/10.1016/j.jhe.2016.05.005

Zhang, D. (2020). The evolution of the wage gap between rural migrants and the urban labour force in Chinese cities. *Australian Journal of Agricultural and Resource Economics*, *64*(1), 55–81. https://doi.org/10.1111/1467-8489.12338

Zhang, H. (2010). The Hukou system's constraints on migrant workers' job mobility in Chinese cities. *China Economic Review*, *21*(1), 51–64. https://doi.org/10.1016/j.chieco.2009.10.002

Zhang, H. (2017). Opportunity or new poverty trap: Rural-urban education disparity and internal migration in China. *China Economic Review*, *44*, 112–124. https://doi.org/10.1016/j.chieco.2017.03.011

Zhang, J., Huang, J., Wang, J., & Guo, L. (2020). Return migration and Hukou registration constraints in Chinese cities. *China Economic Review*, *63*, 101498. https://doi.org/10.1016/j.chieco.2020.101498

Zhang, J., Wang, R., & Lu, C. (2019). A quantitative analysis of Hukou reform in Chinese cities: 2000–2016. *Growth and Change*, *50*(1), 201–221. https://doi.org/10.1111/grow.12284

Zhang, K. H., & Song, S. (2003). Rural–urban migration and urbanization in China: Evidence from time-series and cross-section analyses. *China Economic Review*, *14*(4), 386–400. https://doi.org/10.1016/j.chieco.2003.09.018

Zheng, S., Long, F., Fan, C. C., & Gu, Y. (2009). Urban Villages in China: A 2008 Survey of Migrant Settlements in Beijing. *Eurasian Geography and Economics*, *50*(4), 425–446. https://doi.org/10.2747/1539-7216.50.4.425

Zhou, J., & Tang, S. (2021). Attitudes towards rural migrants and their influence on return migration in China. *Population, Space and Place*, e2509. https://doi.org/10.1002/psp.2509

Zhu, N. (2002). The impacts of income gaps on migration decisions in China. *China Economic Review*, *13*(2), 213–230. https://doi.org/10.1016/S1043-951X(02)00074-3

Zou, W., Liu, L., & Zhuang, Z. (2009). Skill Premium, Biased Technological Change and Income Differences. *China & World Economy*, *17*(6), 64–87. https://doi.org/10.1111/j.1749-124X.2009.01174.x