# Internal Migration and its Socio-Economic Impacts in China: Labour Markets, Financial Decision-Making, and Life Satisfaction

## Shanfei Zhang

Thesis presented for the degree of Doctor of Philosophy in Economics

School of Economics

University of East Anglia

September 2024

# Abstract

This thesis investigates critical socio-economic issues in contemporary China through three distinct studies, each contributing unique insights into migration, mobility, labour markets, and financial decision-making. The first chapter focuses on how internal migration affects local labour market outcomes, particularly wages in urban areas. By employing both linear and nonlinear models, the study identifies a complex relationship where a 10 p.p. increase in migrant share correlates with a 6.78% rise in local wages, especially benefiting skilled workers. Notably, the analysis uncovers a critical threshold of 27.32% migrant share, beyond which the initially negative impact on wages reverses. These results, supported by instrumental variable techniques, provide important implications for policies aimed at optimising the benefits of migration for local economies. The second chapter diverges from the migration theme to focus on household finance. It investigates how family-originated background risks influence the allocation of financial risk assets. Using data from the 2017 China Household Finance Survey (CHFS) and applying Cragg's Double Hurdle Model, this study finds that background risks significantly affect both household participation in risky markets and the extent of risky asset holdings. The findings contribute to a deeper understanding of household financial behaviour under risk and uncertainty. The third chapter explores the reciprocal relationship between migration and life satisfaction in China. Utilising panel data from the China Family Panel Studies (CFPS) from 2012 to 2020, this study examines how life satisfaction influences migration decisions and how these decisions, in turn, affect life satisfaction. The findings reveal that lower life satisfaction increases the likelihood of migration by 29 p.p., while higher life satisfaction discourages mobility. This chapter highlights the significant role of life satisfaction as both a determinant and a consequence of migration decisions, offering nuanced insights into the well-being of migrants in rapidly urbanising regions. These chapters offer a comprehensive analysis of the socio-economic dynamics at play in contemporary China, with particular emphasis on the intersections of migration, labour markets, and financial decision-making.

# Contents

Contents

# List of Figures

# List of Tables

# Acknowledgements

This thesis is the result of a long journey filled with both ups and downs. Throughout this process, I have grown personally and professionally, and I owe much of that growth to the people who have supported me along the way.

First and foremost, I would like to express my deepest gratitude to my family for their unwavering love and support, which have sustained me throughout this journey. Their constant belief in me has been a source of strength during the most challenging moments. I am especially thankful to my parents and grandparents, whose encouragement and reassurance helped me to keep moving forward. Their unconditional support has given me the confidence to pursue my dreams, knowing that I always had them by my side. Above all, I am profoundly grateful to my late grandmother. She was my greatest source of comfort and strength, always believing in my potential even when I doubted myself. You were the most cherished person in my life, and not a day goes by that I don't wish you were here to witness this milestone. In addition to my immediate family, I want to extend my heartfelt thanks to my family in Christ, Dr. Silvia Ferrini and Mr. Arturo Fanciulli, and their daughters. Their kindness, generosity, and moral support have been a blessing to me during this journey. I am truly fortunate to have them as part of my life, and I will always cherish their role in this achievement.

I would like to extend my sincere thanks to Prof. Mich Tvede, who served as my primary supervisor during the first year of my Ph.D. His passion for the subject and insightful guidance during those formative months set the foundation for my research. I am also immensely thankful to Dr. Martin Bruns, who has been my dedicated secondary supervisor from the very beginning of this journey to its completion. His unwavering support and encouragement, even in the most challenging times, have been instrumental in helping me stay focused and resilient. His constructive feedback, patience, and willingness to explore new ideas with me have made this entire process both stimulating and rewarding. I would also like to express my deep appreciation to Dr. Gustavo Fruet Dias, who graciously took over as my primary supervisor during my second year. His expertise and enthusiasm for research have been truly inspiring. He not only guided me through complex theoretical frameworks but also provided

Contents

practical advice that helped me refine my research methodology. His mentorship went beyond academic support; his genuine belief in my abilities and his dedication to my success gave me the confidence to tackle new challenges and embrace the uncertainties inherent in academic research. They accompanied me every step of the way, providing not only their academic wisdom but also their personal support and encouragement. For that, I am profoundly grateful, and I feel truly fortunate to have had the privilege of working with them. Their influence has made a lasting impact on my academic and personal development, and I will carry their lessons with me long into the future.

I would also like to express my heartfelt appreciation to my dear friends from my cohort. Special thanks go to Weijia Shao, Andrea Marietta Leina, Nikita Grabher-Meyer, and Kensley Blaise. The Ph.D. journey is long and challenging, and without their companionship, these past four years would not have been the same. They have become a significant part of my life, and Norwich will never feel the same without them. I would like to thank my dearest friend, Wiktoria Wiatrowska, for always being there for me and offering unwavering support. She reached out to me when I was in my darkest time. From her, I sincerely feel that in life, we may have many friends who add joy and colour, but a true friend who helps in times of need is even more precious.

Furthermore, I extend my deepest gratitude to my colleagues from the School of Economics, particularly Dr. Stefania Sitzia, Dr. Michael Brock, Dr. Rui Silva, and Dr. Andrea Calef. Their invaluable support, insightful academic discussions, and genuine friendship have played a pivotal role throughout my Ph.D. journey. Beyond the academic realm, their camaraderie and warmth made the School of Economics feel like a true community. Whether it was a quick chat in the hallway or a deeper conversation over coffee, their presence has made my time here much more rewarding and enjoyable. I am truly thankful for the friendships we have built, and I look forward to staying connected with them as colleagues and friends in the years to come.

Last but certainly not least, I want to express my deepest and most heartfelt gratitude to my beloved partner, Dr. Valentina Di Gennaro. Her unwavering support and love have been my greatest source of strength throughout this journey, offering not only practical help but also emotional comfort and encouragement when I needed it most. Her belief in me, even during moments of self-doubt, gave me the confidence to persevere. She has been a constant pillar of support, always reminding me of the bigger picture and helping me maintain balance in both my academic and personal life. Words cannot fully capture how grateful I am for her presence in my life; she has truly made this journey more meaningful. I am fortunate beyond measure to walk this path with her by my side.

# Introduction

Migration and its economic and social ramifications are fundamental to understanding labour markets, household decision-making, and individual well-being, particularly in rapidly transforming economies such as China. This thesis examines these complex interrelations across three distinct chapters. While Chapters 1 and 3 focus on the interconnected effects of internal migration on labour markets and life satisfaction, Chapter 2 independently investigates the influence of family background risks on household financial behaviour. Together, these chapters provide a comprehensive and nuanced exploration of the multifaceted impacts of migration and risk on economic outcomes and subjective well-being.

Chapter 1 explores the externalities of internal migration on local labour market outcomes in Chinese cities, with a particular emphasis on wage dynamics. Drawing on both linear and nonlinear models, the analysis reveals a critical threshold in the relationship between migrant share and wage levels. While low levels of migration are associated with wage suppression, particularly for low-skilled workers, the study identifies a turning point beyond which the influx of migrants begins to exert a positive effect on wages, particularly benefiting skilled workers. The study employs instrumental variables to account for potential endogeneity, offering robust insights into the nonlinear labour market adjustments triggered by internal migration.

Chapter 2, in contrast, presents an independent investigation into the effects of family background risks on household financial behaviour, focusing on the allocation of financial risk assets. Employing Cragg's Double Hurdle Model and utilising household-level data from the 2017 China Household Finance Survey, this chapter examines how background risks—including labour income uncertainty, health risks, housing ownership, and business risks—influence household investment decisions. The findings indicate that households with stable income streams are more inclined to engage in riskier financial markets. In contrast, households facing heightened health or housing risks adopt more conservative financial strategies. This chapter contributes to the literature on household finance by highlighting the role of background risks in shaping financial decision-making under uncertainty, independent of migration dynamics.

Chapter 3, closely related to Chapter 1, examines the reciprocal relationship between migration and life satisfaction. While migration is often motivated by economic considerations, this chapter emphasises the role of life satisfaction both as a determinant and an outcome of migration. Using longitudinal data from the China Family Panel Studies (CFPS), the analysis demonstrates that individuals experiencing lower levels of life satisfaction are more likely to engage in mobility and that migration subsequently enhances life satisfaction. Additionally, the chapter addresses return migration, assessing whether individuals who leave major urban centres experience improved life satisfaction upon returning to their places of origin. The combined insights from Chapters 1 and 3 provide a holistic view of migration's dual impact on labour market dynamics and subjective well-being.

While Chapters 1 and 3 are thematically linked through their focus on migration and its effects on labour markets and life satisfaction, Chapter 2 stands as an independent contribution, offering a detailed examination of household financial behaviour in the face of non-financial risks. Together, the chapters in this thesis offer a multidimensional perspective on how migration and background risks shape economic and social outcomes in China. These findings have significant implications for policymakers, particularly in formulating strategies to harness the benefits of migration, mitigate household financial vulnerabilities, and improve the well-being of both migrants and non-migrants.

# Chapter 1

# The Externality Impact of Internal Migration in China: Linear and Nonlinear Approach

This study investigates the impact of internal migration on local labour market outcomes in Chinese cities, focusing on how the influx of migrants influences the wages of native workers. By employing both linear and non-linear models, the analysis reveals a complex relationship between migrant share and local wages. The linear results indicate that a 10 p.p. increase in migrant share has an effect with a 6.78% rise in local hourly wages, particularly benefiting skilled workers. However, the non-linear analysis highlights a turning point. While there is an initial negative effect on wages when the migrant share is low, this trend reverses once the migrant share exceeds 27.32%, resulting in wage increases for native workers. The study employs instrumental variables to address potential endogeneity, thereby enhancing the robustness of the findings. These insights contribute to a deeper understanding of the complex dynamics of internal migration and its influence on wage structures in urban China, offering valuable implications for policy interventions aimed at optimising the benefits of migration.

## 1.1   Introduction

Before 1980, China operated under a planned economy. Prior to the late 1980s implementation of the opening-up policy, individuals were mandated to live and work solely within the jurisdiction defined by their 'Hukou'[1] (Davin, 1998). The primary differential between internal migrants and native residents is the extent of benefits they can obtain from the city they inhabit, such as public welfare and educational opportunities (Démurger et al., 2009). This framework has imposed stringent restrictions on population movement for decades. Post-1990s, however, China witnessed substantial internal labour migration, catalysing rapid industrialisation and urbanisation, and engendering significant population migration and labour mobility (Hering and Poncet, 2010). Empirical evidence substantiates that domestic population mobility constitutes the principal driver of urban population growth (Chen et al., 2016). According to the 2005 national sample survey, which covered 1% of the population, the floating population[2] who had resided outside their registered household location for more than six months numbered 147 million, representing over 11% of the total population. Individuals aged 16 to 60 constituted 80.2% of the floating population. Data from the sixth national census in 2010 indicated that the floating population had surged to 221 million, marking a 116.7% increase from 2000. Labour migration exerts distinct effects on labour markets in host regions.

In the process of urbanisation, human capital differences supplant traditional industry barriers as the primary determinant in securing employment opportunities (Sullivan, 2010; Xuguang and Jiajia, 2022). The theory of competitive labour markets posits that foreign and local labours are entirely substitutable, highlighting potential negative impacts on local workers (Borjas, 2003). While low-skilled local workers may face competitive pressures from immigrants, the influx of immigrants can also trigger the relocation of locals who feel threatened by competition, potentially increasing the demand within the local labour market and thereby mitigating some of these impacts (Mouw, 2016). Furthermore, other research suggests that foreign labour can enhance labour productivity and foster job reclassification, ultimately boosting employment rates and wages within the local labour market (Ottaviano and Peri, 2012). Regarding the relationship between labour mobility and income gap, while international studies predominantly focus on

---

[1] The 'Hukou' system, established in 1958 and discontinued in 1979, functioned as a migration control mechanism, classifying individuals as either rural or urban and assigning them to specific geographic locales. Movement between these locales was strictly regulated, and residents were denied access to employment, public services, education, healthcare, and food outside their designated areas.

[2] The floating population comprises all individuals without local residency rights, including both migrant workers and their families (non-working individuals), originating from either rural or urban areas. It excludes those who have obtained a change in their household registration.

the effects of transnational immigration on labour market inequalities in host countries, research within China primarily examines the factors contributing to income disparities across local and foreign, regional, and urban-rural labour markets. Card (2009) finds that immigration bears no significant correlation with the wage disparities between high-skilled and low-skilled workers. Similarly, Blau and Kahn (2015) find that although immigration may alter the skill composition of the workforce, its overall impact on local wage structures is negligible. Furthermore, Lewis and Peri (2015) conclude that immigration exerts a favourable influence on the wages of most native-born workers in industrialised nations, according to their comprehensive review of the extensive and frequently contested literature regarding the impacts of immigration on local economies.

As the migrant population grows both in number and as a proportion of the local resident population, and despite their concentration in specific segments of local labour markets, concerns persist regarding their potential negative effects on the labour market outcomes of local residents. Does the influx of immigrants indeed impact the interests of local residents and affect the wage rates of the local labour force? This question is pivotal for scholars and government officials alike, bearing significant implications for regional economic development and the potential future abolition of the enduring "Hukou" system. These effects hinge on the degree of substitutability or complementarity between immigrant and native workers, which in turn depends on the characteristics of the native workers. Moreover, when faced with immigration, native workers may choose to relocate either spatially or occupationally in response to heightened competition, and this mobility could alleviate the potential adverse impacts of immigration. Acemoglu and Angrist (2000) believes labour inflow externalities are categorised into non-monetary and monetary externalities, realised through knowledge exchange and increases in physical capital, respectively. Moreover, Labour inflow fosters economic development in the destination area via the agglomeration effect of human capital and enhances the wage levels of residents in these areas. Therefore, in the process of China's urbanisation development, it is necessary to deeply examine the impact of floating population on local residents based on the perspective of heterogeneous labour mobility with different human capital endowments.

This paper evaluates the impact of migrants on the wages of local residents. The impact of immigrants on the wage rate of local labor is examined by linear and nonlinear models. Our analysis demonstrates a significant negative linear impact of immigrants on local labour wages. Specifically, for every 10 p.p. increase in the immigrant population, there is a corresponding 6.78% decrease in the wages of local workers. However, nonlinear estimations indicate that as the immigrant proportion

approaches a critical threshold, the negative impact on local wages diminishes, as evidenced by the quadratic term in our model. Notably, when the proportion of the foreign population exceeds 27.32%, the effect on local labour shifts from negative to positive, with the positive marginal effect increasing progressively. Furthermore, our findings show that migrants have a significantly positive impact on the wages of local residents across all skill levels, with more pronounced benefits for skilled workers.

This chapter delves deeper into the influence of the foreign population on the wage rates of local labour in Chinese cities, utilising more recent datasets to provide a realistic portrayal of immigration issues. This facilitates a quantitative evaluation of current government policies and allows for the formulation of specific recommendations. Previous research often overlooked important differences within the labour force, typically basing conclusions on generalised assumptions about the 'substitution' or 'complementarity' relationships between migrant and local workers. These studies ignore variations in skill levels by migrants versus natives. As a result, they failed to capture the nuanced interactions between different segments of the migrant and local labour forces, leading to an underappreciation of how these groups influence each other's wages and employment outcomes.

The following content of the article is arranged as follows: Section 1.2 is the literature review. Section 1.3 introduces the database, and main variable relations, Section 1.4 shows the empirical models. Sections 1.5 and 1.6 parts report the results. Section 1.8 is the conclusion.

## 1.2 Literature Review

Two main types of research are closely related to this article: the impact of labour mobility on the local job market and the consequences of labour mobility on income inequality. Theoretically, the impact of China's labour inflow on the local labour market mirrors the impact of transnational immigration on the host country's labour market, though the conclusions are not uniform. While most literature focuses on international immigration or employs foreign immigration as a research model, relatively few studies address internal cases. Furthermore, the relationship between labour mobility and income disparities varies between international and domestic contexts. International studies predominantly examine the effects of transnational immigration on labour market inequality in host countries. Conversely, domestic research primarily investigates the drivers of income gaps between locals and foreigners, across regions, and between urban and rural areas within labour markets.

There are three main perspectives on the impact of labour mobility on the local job market. The first perspective posits that immigration significantly negatively

impacts local labour wages. Immigration increases the local labour supply, causing wage rates to fall as the labour market reaches a new equilibrium, creating competition with local labour and reducing wages within skill groups (Borjas et al., 1997; Borjas, 2003; Aydemir and Borjas, 2006; Cohen-Goldner and Paserman, 2011). The second perspective argues that while the local low-skilled labour force may experience competitive pressures from immigrants, the arrival of these immigrants could prompt the outflow of local residents who feel threatened by this competition, thereby enhancing the demand curve in the local labour market, so that immigration has a minor or zero impact on local labour wages (Card, 1990; Friedberg, 2001; Mouw, 2016). The third perspective suggests immigration positively impacts local labour wages, indicating labour is not homogenous. Heterogeneous immigrants and local labour are often not substitutes and may be complementary, enhancing productivity and increasing wage rates. Studies show immigration prompts firms to change production techniques, expanding industries, particularly low-skilled labour industries (Altonji and Card, 1991; Hanson and Slaughter, 2002; Card and Lewis, 2007; Lewis, 2011). Immigration promotes specialised labour division, with immigrants and locals choosing jobs based on comparative advantages, complementing each other (Peri and Sparber, 2009). Considering labour heterogeneity, immigrants and local labor can be complementary within different groups (Peri, 2007; Ortega and Verdugo, 2011; Ottaviano and Peri, 2012; Combes et al., 2015).

From the perspective of income inequality or disparity, the association between migration and local labours is not clear either. Card (2009) posit that immigration impacts only natives who have comparable skills and concluded that immigration does not influence the wage gap between high-skilled and low-skilled labour. Similarly, Blau and Kahn (2015) observe that although immigration alters the skill composition of the local workforce, its impact remains minimal. Other studies suggest that labour mobility contributes to widening income inequality in destination areas. Dustmann et al. (2013) report that immigration reduces wages for locals in the lowest 20% of the wage distribution but raises wages for those in the upper half. Combes et al. (2020) assess the effects of labour mobility and urbanisation in China, finding that high-skilled locals gain the most from agglomeration and labour mobility at the city level. Additionally, the influx of immigrants can stimulate foreign capital inflow or local investment, creating complementary employment opportunities (Chiswick et al., 1992; Leamer and Levinsohn, 1995; Hamermesh, 1996). The consumption demands of immigrants enhance local market consumption (Orrenius and Zavodny, 2008). The local labour supply adjusts as some local workers exit under competitive pressure, thereby alleviating local market tensions (Card, 2001).

Moreover, inter-regional trade helps integrate the labor market, mitigating the negative impacts of immigration (Borjas et al., 1997).

Although competitive market theory intuitively offers a framework to assess the impact of immigration on native workers with comparable skills, however, understanding how immigration affects native wage rates is no longer a simple static analysis under the assumption of labour homogeneity. Notably, there exists a cross-skill complementarity between the migrant and local workforce, which includes interactions between low-skilled migrants and high-skilled locals, as well as high-skilled migrants and low-skilled locals. Specifically, due to occupational skill segmentation, low-skilled migrants predominantly undertake jobs that local workers are reluctant to perform, especially roles characterised by low technical requirements or substantial physical labour (Knight et al., 2013; Peri and Sparber, 2011). While the influx of such labour preserves low-skilled economic activities, it does not fully eliminate the competitive dynamics between low-skilled migrants and local low-skilled labour. Moreover, high-income groups may choose to outsource certain services to extend their productive hours. Cortes and Tessada (2011) observe that the introduction of low-skilled labour is linked to a significant increase in working hours among highly educated women. Therefore, the extent of complementarity could correlate with the opportunity cost of local labour, with higher local wage levels potentially translating into greater benefits from the influx of low-skilled labour. When considering high-skilled migrants and low-skilled locals, the interactions predominantly manifest through production and consumption channels. On the production front, high-skilled migrants can stimulate sectors like information technology, indirectly increasing the demand for low-wage support services. On the consumption side, the arrival of high-skilled labour enhances overall expenditure on non-traded goods and services, elevating wage expectations among low-skilled service workers (Autor and Dorn, 2013). Consequently, the influx of high-skilled labour tends to improve the labour market performance of local low-wage workers. Thus, as Borjas and Van Ours (2010); Lewis and Peri (2015) articulate, skill interactions must be considered and integrated when evaluating the wage impacts (externalities) of immigration.

Therefore, how immigrants affect the wage rate of the local labour force is still a question that requires empirical research in China. I first employ OLS estimation to investigate the linear relationship between the migrant share and native workers' wages. Subsequently, I employ 2SLS estimation to address the potential endogeneity of migrant share, ensuring that the results are more robust and reliable. Secondly, based on the OLS estimation, I use the quadratic form model to explore whether there is a nonlinear relation, thereby evaluating the impact of different immigration ratios on

local wages from a quantitative perspective. Lastly, I address the question of whether there is a cross-skill complementarity between the labour force (migrant and local).

## 1.3 Data and Instruments Description

### 1.3.1 Data Sources

To test whether the immigrant share significantly affects the wage rate of local workers and cross-skill complementarity between the migrant and local workforce, the data used for this chapter are drawn from three main sources, first, the Chinese Household Finance Survey (CHFS)[3] conducted by the Survey and Research Center for China Household Finance at the Southwestern University of Finance and Economics. Second, migrant share in cities extracted from the Wind [4]. Third is city-level data compiled from the 2017 China City Statistical (CCS) Yearbook published by the National Bureau of Statistics of China (NBS). Additional data for instruments come from historical maps and from Au and Henderson (2006); Jia (2014); Combes et al. (2015).

To tailor the raw data to our research objectives, we implemented several processing steps. The sample exclusively comprises registered urban residents (i.e., holders of urban Hukou), thereby focusing our analysis on wage disparities among native urban residents. The regression sample is confined to the working-age population, specifically males aged 16 to 60 years and females aged 16 to 55 years who reported working at least part of the year and earning a positive wage. We excluded students currently enrolled in educational institutions, individuals without employment status, and non-earning segments. This chapter also controls for key factors such as workers' education levels, potential work experience, and marital status, ensuring that these variables are properly accounted for in the analysis. The refined sample of observations is 8,066. Different occupations may affect the rate of return of wages. The questionnaire asks about the individual's work occupation [5], therefore, we could add seven dummy

---

[3]Due to traceable observations during four consecutive waves of this survey are very small, which limits our econometric model to cross-sectional. The survey sample covers 29 provinces (autonomous regions and municipalities directly under the Central Government), 355 counties (districts, county-level cities), and 1428 village (residential) committees, and the sample size covers 40011 households.

[4]A database that pairs over 1.3 million macroeconomic and industry time series with data analysis tools to give financial professionals the most comprehensive insights into China's economy

[5]The standard stipulates the classification structure, categories, and codes of Chinese occupations, and is applicable to various censuses, survey statistics, administrative management, and domestic and foreign information exchanges. Occupation classification is based on the identity and nature of the work performed by the working population. According to regulations, the national occupational classification is divided into 8 major categories, namely the first category: heads of state agencies, party organisations, enterprises, and institutions; the second category: professional and technical personnel; the third category: office workers and related personnel; The fourth category: commercial

variables in order to fix the influence of occupations. The dependent variable is defined as the declared hourly income from wage employment. This encompasses basic salary, various bonuses, allowances, and subsidies (including those for housing or medical purposes), overtime wages, and other income derived from the employment unit.

The share of migrants within this framework is quantified based on their proportion in city employment, calculated using the Wind dataset. Migrants are characterised as working individuals aged 16–60 who do not reside in their county of Hukou registration, including those from both rural and other urban areas. City-level data are sourced from the China City Statistical Yearbook 2017. China's administrative hierarchy consists of four tiers. As of 2017, the mainland was organised into 32 province-level regions (excluding the two Special Administrative Regions, Hong Kong and Macao), and comprised 286 prefecture-level cities (National Bureau of Statistics, 2018). Prefecture-level cities in China generally encompass the central city, its suburban districts, and several rural counties that fall under the city government's administration. Statistics on employment and area are provided for two distinct levels: the entire prefecture and the metropolitan area, which includes only the city and its suburban districts. For the purposes of this analysis, all city-level variables pertain to the metropolitan area.

Tables 1.1 and 1.2 presents the symbols for each variable, their definitions, and the sources from which the data for each variable is obtained. Table 1.3 reports each variable's number of observations, the mean value, standard deviation, and minimum & maximum value.

Table 1.1: Definition of Variable

| Variable | Description | Data Source |
|---|---|---|
| lnwage | natural logarithm of the native individual | CHFS |
| ms | $migrantshare = migrantpopulation/totalpopulation$ | Wind |
| educ | education years at individual level | CHFS |
| male | dummy variable, male $= 1$, $= 0$ otherwise | CHFS |
| exp | Potential working experience years, calculated by *individual's age - year of schooling - 6* | CHFS |
| exp_sq | the quadratic working experience | CHFS |
| marriage | dummy variable, $= 1$ if ever married; $= 0$ otherwise | CHFS |
| Occ1 | The leader of the party and state organs, mass organisations, social organizations, enterprises, and public institutions | CHFS |

Note: Gender, working experience, and its square term, marriage, and education are at the individual level; other variables are at the city level.

and service personnel; the fifth category: agricultural, forestry, animal husbandry, fishery, and water conservancy production personnel; the sixth category: production and transportation equipment operators and related personnel; the seventh category: military personnel; the eighth category: others.

Table 1.2: Continued Table 1.1

| Occ2 | Technicians and professionals | CHFS |
|---|---|---|
| Occ3 | Clerks or related staffs | CHFS |
| Occ4 | Social production and life service workers | CHFS |
| Occ5 | Production operators and support staff of the industry of agriculture, forestry, husbandry, fishery | CHFS |
| Occ6 | Wholesale and retail | CHFS |
| Occ7 | Transportation, storage, logistics and mail | CHFS |
| Occ8 | Others without clear classification | CHFS |
| lnemp_den | natural logarithm of the employment density ($10000\ employment/km^2$) | CCS |
| lnpubexp | natural logarithm of public expenditure | CCS |
| lnexpst | natural logarithm of public expenditure in science and technology | CCS |
| lnexpedu | natural logarithm of public expenditure in education | CCS |
| lnfixedasset | natural logarithm of total fixed asset | CCS |

Note: Gender, working experience, and its square term, marriage, and education are at the individual level; other variables are at the city level.

Table 1.3: Data Description

| Variable | Obs | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|---|
| lnwage | 12,128 | 5.013 | 0.686 | 2.303 | 9.554 |
| ms | 12,128 | 0.141 | 0.14 | 0 | 0.774 |
| ms_sq | 12,128 | 0.039 | 0.090 | 0 | 0.599 |
| educ | 12,128 | 11.358 | 3.656 | 0 | 22 |
| male | 12,128 | 0.602 | 0.490 | 0 | 1 |
| exp | 12,128 | 21.464 | 12.395 | 0 | 54 |
| exp_sq | 12,128 | 614.334 | 555.216 | 0 | 2916 |
| marriage | 12,128 | 0.811 | 0.391 | 0 | 1 |
| lnemp_den | 12,128 | 6.730 | 0.767 | 2.628 | 8.373 |
| lnpubexp | 12,128 | 15.403 | 0.692 | 12.400 | 17.556 |
| lnexpst | 12,128 | 11.120 | 1.491 | 6.624 | 15.211 |
| lnexpedu | 12,128 | 13.638 | 0.690 | 10.555 | 15.238 |
| lnfixedasset | 12,128 | 16.866 | 0.866 | 13.545 | 18.241 |
| Occ1 | 12,128 | 0.031 | 0.173 | 0 | 1 |
| Occ2 | 12,128 | 0.216 | 0.412 | 0 | 1 |
| Occ3 | 12,128 | 0.224 | 0.417 | 0 | 1 |
| Occ4 | 12,128 | 0.216 | 0.412 | 0 | 1 |
| Occ5 | 12,128 | 0.014 | 0.118 | 0 | 1 |
| Occ6 | 12,128 | 0.200 | 0.400 | 0 | 1 |
| Occ7 | 12,128 | 0.002 | 0.046 | 0 | 1 |

Note: Gender, working experience and its square term, marriage, and education are at the individual level. Other variables are at the city level.

## 1.4　Methodology

This section outlines the methodology for simultaneously identifying substitutability and complementarity, along with the agglomeration effects of migrants. It also addresses potential endogeneity biases that may result from omitted variables and reverse causality.

The literature on immigration highlights that migrants can influence the labour market outcomes of native workers either positively or negatively. A positive impact occurs when the skills of migrants complement those of native workers, while a negative impact arises when migrant skills substitute for those of natives, see discussion in Section 1.2. Equation (1.1) represents that the share of migrants impacts natives' wages which is regressed on individual and city control variables.

$$lnwage_{i,j}^{Native} = \beta_0 + \beta_1 MS_j + \boldsymbol{\alpha X_{i,j}} + \boldsymbol{\kappa City_j} + \sum_{i=1}^{7} \boldsymbol{\gamma Occ_i} + \epsilon_{i,j} \qquad (1.1)$$

where the subscripts $i$ and $c$ respectively represent the individual $i$ and its location of the city, and $\sum_{i=1}^{7} \boldsymbol{\gamma Occ_i}$ is a vector that refers to seven occupational dummy variables, and $\epsilon_{i,j}$ represents the random disturbance term. The explained variable of the model $lnwage_{i,j}^{Native}$ represents the natural logarithm of hourly wage income of the natives. $X_{i,j}$ is a vector of individual features that may affect wages, including gender, marital status, potential work experience, and its square term. We do not include age in the personal characteristics because the variable of potential work experience has been included in the regression, which is derived from *age – year of schooling - 6*. If both age and work experience are included in the regression, it brings the problem of perfect collinearity. Urban features that may affect employment are included in the $City_j$ vector, including the natural logarithm of the average fixed asset, the average of the government public expenditure, the public expenditure in education, science and technology, and employment density. The core explanatory variable of the model is $MS$, which is given by $MS_j = \frac{Migrants_j}{Migrants_j + Natives_j}$. $Migrants_j$ and $Natives_j$ respectively represent the number of immigrants and the number of local workers in the $j$ city. A positive $\beta_1$ demonstrates a scenario in which complementarities between natives and migrants lead to an increase in the wages of native workers as the local migrant share rises. Lewis and Peri (2015); Combes et al. (2015) highlight that in this case, $\beta_1$ only reflects the partial effect of migrants, assessed within the specific context of the natives' chosen occupation and location.

As migrants influence native wages, this interaction should subsequently affect the occupational and location choices of natives, suggesting a feedback loop. When local economic factors are generated from agglomeration economies, these economies influence local productivity through technological spillovers and pecuniary

economies, which improve access to markets for final goods and intermediate inputs, thereby affecting local nominal wages. This may raise a concern that impacts natives' wages in ways not fully captured by $\beta_1$. However, in the context of China, where the occupational choices of local residents and migrants are often segregated and mobility is limited, Combes et al. (2015) argue that this feedback should not significantly alter the distinction between partial and total effects of migrants. Also, in Equation (1.1), we define the employment density in city $j$ as $Emp\_Den_j^{Total} = \frac{Emp_j^{Total}}{Area\_Size_j}$, where $Emp_j$ is the number of total employment and $Area\_Size_j$ is the size of $city_j$. We can further deduce the variables as follows:

$$Emp\_Den_j = Emp\_Den_j^{Native}\left(1 + \frac{Emp_j^{Migrant}}{Emp_j^{Native}}\right) = \frac{Emp\_Den_j^{Native}}{1 - MS_j} \qquad (1.2)$$

where $Emp_j^{Native}$ and $Emp_j^{Migrant}$ are the numbers of native and migrant employed in the city $j$, $Emp\_Den_j^{Native}$ is the naive employment density in city $j$.

Therefore, we can rewrite Equation (1.1) as,

$$lnwage_{i,j}^{Native} = \beta_0 + \beta_1 MS_j + \tau_1 \frac{Emp\_Den_j^{Native}}{1 - MS_j} + \boldsymbol{\alpha X_{i,j}}$$
$$+ \boldsymbol{\kappa City_j} + \sum_{i=1}^{7} \boldsymbol{\gamma Occ_i} + \epsilon_{i,j} \qquad (1.3)$$

According to Combes et al. (2013) assumptions and the empirical results from Combes et al. (2015), we can simplify Equation (1.3) to Equation (1.4), because when the agglomeration effects of local residents and immigrants in cities are comparable. This implies that when considering a city's total employment density, there is no need to differentiate between the effects of local population density and immigrant density. Consequently, the substitution or complementary effects of immigrants can be captured by a distinct immigrant ratio parameter, allowing the direct impact of the immigrant ratio on wages to be analysed independently of the total density impact.

$$lnwage_{i,j}^{Native} = \beta_0 + \beta_1 MS_j + \tau_1 lnemp\_den_j^{Total} + \boldsymbol{\alpha X_{i,j}}$$
$$+ \boldsymbol{\kappa City_j} + \sum_{i=1}^{7} \boldsymbol{\gamma Occ_i} + \epsilon_{i,j} \qquad (1.4)$$

Due to constraints related to the cross-sectional data, we are unable to analyse the impact of the migrant share on local labour wage rates over time. Nevertheless, to address this within a static time framework, we incorporate a quadratic term of the proportion of immigrants into our analysis. This methodological approach allows us

to explore the specific thresholds at which the proportion of the migrant population in Chinese cities may exhibit either crowding-out effects, where migrants displace local workers, or complementary effects, where migrants and local workers enhance each other's productivity. When the coefficient of migrant share (MS) is negative, but the coefficient of the quadratic term of $MS$ is positive, which means that the equation with the quadratic term has a U-shaped relationship like a parabola. It shows that at the turning point $(MS^*)$ the marginal influence of MS on local labours' wage rate is zero, before this point $(MS < MS^*)$, the influence of MS on the dependent variable is negative; and after this point, the influence of internal migrants on locals is positive. This quantitative estimation is of policy significance in controlling the city size.

$$lnwage_{i,j}^{Native} = \beta_0 + \beta_1 MS_j + \beta_2 MS_j^2 + \tau_1 lnemp\_den_j^{Total}$$
$$+ \boldsymbol{\alpha X_{i,j}} + \boldsymbol{\kappa City_j} + \sum_{i=1}^{7} \boldsymbol{\gamma Occ_i} + \epsilon_{i,j} \qquad (1.5)$$

Lastly, I add robust standard error and cluster robust standard error to solve the potential problems arising from heteroskedasticity and correlated error terms within a cluster. Because the homoscedasticity assumption is generally not satisfied in cross-sectional research. If heteroscedasticity exists, the ordinary standard error that assumes homoskedasticity is not consistent (Woodridge, 2010).

### 1.4.1 Addressing Endogeneity

Stemming from reverse causality between the migrant share and local residents' wage levels may affect the precision of our analysis, which leads to biased or inconsistent estimates if not properly addressed. The second concern relates to employment density, which, although not a core explanatory variable in our regression, may also exhibit reverse causality through the migrant share channel. The rationale for this is that first, immigrants act both as responders to and determinants of wage levels; high wages attract immigrants, while immigrants influence labour markets and economic activities, thereby impacting wages. This establishes a bidirectional causal relationship between the proportion of immigrants and wage levels. Moreover, while it is commonly assumed that immigrants increase the employment density of cities, high wage levels can further elevate employment density by attracting both more immigrants and local labour. This interdependence suggests that although employment density may influence wage levels, it can also be an outcome of high wage levels. To mitigate the issue of endogeneity, we employ an instrumental variable approach, utilising several sets of instruments to source exogenous variations for the variables suspected of endogeneity. The method allows us to isolate the causal effects

of the endogenous variables by introducing external variables that are correlated with the endogenous variables but uncorrelated with the error terms in the regression model. By doing so, we aim to ensure the robustness and validity of our findings.

We adopt an instrumental variable approach, utilising several sets of instruments to source exogenous variations for the suspected endogenous variables. This method helps to isolate and assess the true causal relationships by leveraging variables that are correlated with the endogenous predictors but uncorrelated with the error terms in the regression model. This strategic choice enables a more precise estimation of the effects, mitigating potential biases caused by endogeneity.

First, although acquiring historical population data for different cities can be quite challenging, some scholars have argued that geographic data can serve as effective instruments. Jia (2014) uses the classification of whether a city served as a treaty port after the Opium War as an instrumental variable. The rationale behind this choice is that proximity to historically significant cities can be a substitute for being located there, given that interactions remain possible within reasonable distances. Similarly, I have considered an overall peripherality index, which encompasses the average distance of any city to all cities, both historic and non-historic, as outlined by (Combes et al., 2015). Since treaty ports are largely synonymous with seaports, I introduce the average distance (peripherality) to seaports as an instrument. However, I enhance this measure in my research, as it may underestimate the geographic advantage of cities in close proximity to seaports. To address this limitation, I have introduced the concept of minimum distance to the nearest seaport. To obtain this variable, I have undertaken a three-step process: Firstly, calculate spherical distances between the target city and seaport cities using (Vincenty, 1975). Second, computing the minimum value among these distances. Third, transform these values into natural logarithms to derive the instrument variable. Economic geography theory suggests that peripherality instruments are not weighted by population or employment, ensuring their correlation with instrumented variables while reducing the likelihood of strong association with current wage shocks.

The third instrument I incorporate pertains to the industry structure of cities. Given the economic development over the past decades, one key factor influencing labours' wages is the industrial composition of a city. For instance, international metropolises like Shanghai exhibit a significant share of financial services, resulting in higher average wages and attracting a large number of migrants. To capture this aspect, I use the ratio of GDP in 1990 generated by the service sector to the combined GDP of the other two sectors as a measure of the industry structure. With a lag of over two decades spanning a period of significant reforms in China, the historical

industrial composition is unlikely to exhibit a strong correlation with contemporary wage shocks.

Therefore, the TSLS of Equation (1.4) is rewritten in follow Equations (1.6) to (1.8). Equations (1.6) and (1.7) are the First Stage of Instrumental Variable Process to Equation (1.8). **Control_Variables** is a vector of control variables including individual, city-level, and occupational, see Section 1.4.

$$\widehat{MS_j} = \omega_0 + \omega_1 Structure\_90_j + \omega_2 lmin\_dis_j + \boldsymbol{\omega Control\_Variables} + \vartheta_{i,j} \quad (1.6)$$

$$\widehat{lnemp\_den}_j^{Total} = \upsilon_0 + \upsilon_1 Structure\_90_j + \upsilon_2 lmin\_dis_j \\ + \boldsymbol{\upsilon Control\_Variables} + \theta_{i,j} \quad (1.7)$$

$$lnwage_{i,j}^{Native} = \beta_0 + \beta_1 \widehat{MS_j} + \tau_1 \widehat{lnemp\_den}_j^{Total} + \boldsymbol{\alpha X_{i,j}} + \boldsymbol{\kappa City_j} \\ + \sum_{i=1}^{7} \boldsymbol{\gamma Occ_i} + \epsilon_{i,j} \quad (1.8)$$

Moreover, the TSLS of Equation (1.5) is rewritten in follow Equations (1.9) to (1.12). Equations (1.9) to (1.11) are the First Stage of Instrumental Variable Process to Equation (1.12).

$$\widehat{MS\_sq_j} = \pi_0 + \pi_1 Structure\_90_j + \pi_2 Structure\_90\_sq_j + \pi_3 lmin\_dis_j \\ + \pi_4 lmin\_dis\_sq_j + \pi_5 historic + \boldsymbol{\pi Control\_Variables} + \eta_{i,j} \quad (1.9)$$

$$\widehat{MS_j} = \lambda_0 + \lambda_1 Structure\_90_j + \lambda_2 Structure\_90\_sq_j + \lambda_3 lmin\_dis_j \\ + \lambda_4 lmin\_dis\_sq_j + \lambda_5 historic + \boldsymbol{\lambda Control\_Variables} + \zeta_{i,j} \quad (1.10)$$

$$\widehat{lnemp\_den}_j^{Total} = \mu_0 + \mu_1 Structure\_90_j + \mu_2 Structure\_90\_sq_j + \mu_3 lmin\_dis_j \\ + \mu_4 lmin\_dis\_sq_j + \mu_5 historic + \boldsymbol{\mu Control\_Variables} + \varepsilon_{i,j} \quad (1.11)$$

$$lnwage_{i,j}^{Native} = \beta_0 + \beta_1 \widehat{MS_j} + \beta_2 \widehat{MS_j^2} + \tau_1 \widehat{lnemp\_den}_j^{Total} + \boldsymbol{\alpha X_{i,j}} + \boldsymbol{\kappa City_j} \\ + \sum_{i=1}^{7} \boldsymbol{\gamma Occ_i} + \epsilon_{i,j} \quad (1.12)$$

## 1.5 Results

Our baseline Ordinary Least Squares (OLS) estimation results are detailed in Table 1.4. The first column outlines results that incorporate only city-level control variables and occupational effects. Subsequent columns (2) to (5) sequentially introduce individual-level characteristics such as education, gender, work experience and its square term, and marital status. Incorporating individual-level characteristics further solidifies the model and its explanatory power, as evidenced by an increase in the $R^2$ value from 0.159 to 0.266, maintaining its significance at the 1% level. Compared to the initial column, the regression coefficients for the proportion of immigrants increase from 0.397 to 0.446, which is statistically significant at the 1% level. Specifically, holding other variables constant, a 10 p.p. increase in the proportion of immigrants is associated with a 4.46% increase in the hourly wages of local labour. This result indicates a positive relationship between the influx of migrants and wage levels in the local labour market. This improvement in model robustness underscores the importance of including a comprehensive range of factors when analysing the impact of immigration on wage dynamics.

Table 1.4: OLS Estimation Results

|  | (1) lnwage | (2) lnwage | (3) lnwage | (4) lnwage | (5) lnwage |
|---|---|---|---|---|---|
| ms | 0.397*** | 0.423*** | 0.427*** | 0.456*** | 0.446*** |
|  | (0.066) | (0.063) | (0.063) | (0.062) | (0.062) |
| lnemp_den | -0.041*** | -0.048*** | -0.045*** | -0.049*** | -0.049*** |
|  | (0.011) | (0.010) | (0.010) | (0.010) | (0.010) |
| educ |  | 0.052*** | 0.052*** | 0.066*** | 0.065*** |
|  |  | (0.002) | (0.002) | (0.002) | (0.002) |
| male |  |  | 0.179*** | 0.176*** | 0.178*** |
|  |  |  | (0.013) | (0.013) | (0.013) |
| exp |  |  |  | 0.033*** | 0.029*** |
|  |  |  |  | (0.002) | (0.002) |
| exp_sq |  |  |  | -0.001*** | -0.001*** |
|  |  |  |  | (0.000) | (0.000) |
| marriage |  |  |  |  | 0.067*** |
|  |  |  |  |  | (0.023) |
| Constant | 3.460*** | 3.465*** | 3.358*** | 3.217*** | 3.223*** |
|  | (0.255) | (0.246) | (0.243) | (0.239) | (0.239) |
| City Control | YES | YES | YES | YES | YES |
| Occupation Control | YES | YES | YES | YES | YES |
| $N$ | 9,086 | 9,086 | 9,086 | 9,086 | 9,086 |
| $R^2$ | 0.159 | 0.219 | 0.235 | 0.265 | 0.266 |

Note: Robust standard errors in parentheses: ***Significant at 1%; **Significant at 5%; *Significant at 10%. The full table result is in Table 1.10.

In cross-sectional data analysis, heteroskedasticity is a prevalent issue. To address this, I utilised the White test to confirm the presence of heteroskedasticity in the

OLS estimations[6]. These findings affirm the existence of heteroskedasticity within the estimation. Moreover, the regression analysis highlights a pattern where group observations from the same city exhibit identical migrant share (MS) values, suggesting potential correlation within the same clusters, while observations from different clusters remain independent. To accommodate these complexities, I report the OLS results with adjusted standard errors in Table 1.5, taking these intra-cluster correlations into account for a more precise interpretation of the data. In columns (2) and (3), the standard errors, robust and clustered respectively, are recalibrated and show an increase from 0.062 to 0.133. Despite these adjustments, our results remain stable at 0.446 and are highly significant at the 1% level. This consistency underscores the robustness of our findings despite the heteroskedastic nature of the data.

Table 1.5: Ordinary and Adjusted OLS Estimation Results

| | (1) | (2) | (3) |
| | OLS | OLS_R | OLS_C_R |
| VARIABLES | lnwage | lnwage | lnwage |
|---|---|---|---|
| ms | 0.446*** | 0.446*** | 0.446*** |
| | (0.062) | (0.064) | (0.133) |
| lnemp_den | -0.049*** | -0.049*** | -0.049** |
| | (0.010) | (0.010) | (0.020) |
| educ | 0.065*** | 0.065*** | 0.065*** |
| | (0.002) | (0.003) | (0.004) |
| male | 0.178*** | 0.178*** | 0.178*** |
| | (0.013) | (0.013) | (0.013) |
| exp | 0.029*** | 0.029*** | 0.029*** |
| | (0.002) | (0.003) | (0.003) |
| exp_sq | -0.001*** | -0.001*** | -0.001*** |
| | (0.000) | (0.000) | (0.000) |
| marriage | 0.067*** | 0.067*** | 0.067*** |
| | (0.023) | (0.024) | (0.022) |
| Constant | 3.223*** | 3.223*** | 3.223*** |
| | (0.239) | (0.236) | (0.428) |
| City Control | YES | YES | YES |
| Occupation Control | YES | YES | YES |
| N | 9,086 | 9,086 | 9,086 |
| $R^2$ | 0.266 | 0.266 | 0.266 |

Note: Robust standard errors in parentheses: ***Significant at 1%; **Significant at 5%; *Significant at 10%. The full table result is in Table 1.11.

Given the above analysis, in subsequent results reports, I will solely present the OLS estimation and the Two Stage Least Squares (TSLS) estimation, both with robust cluster standard errors.

As we discussed in Section 1.4.1, our first main results are provided in Table 1.6, which displays both OLS and IV estimations, columns (2) and (3) show the first

---

[6]The results of the test yield a chi-square value of $\chi^2(154) = 310.97$, with a significance level of $Prob > \chi^2 = 0.000$.

stage results; and column (4) presents the second step and its specification. The disparities observed between columns (1) and (4) can be attributed to the correction of an underestimation in the OLS results, arising from issues of reverse causality and measurement errors, which are addressed upon the implementation of the instrumental variable (IV) method. The IV model specification has been carefully structured to address under-identification and weak identification issues. As outlined in our methodology, we recognise that migrant share and employment density may exhibit reverse causality with the native labour wage rate. We have utilised two instrumental variables corresponding to these two endogenous variables; this configuration leads to the model being exactly identified, thus eliminating the need to perform the over-identification test (Sargan test). Furthermore, the Chi-sq(3) p-value of the Anderson LM statistic is significantly less than 0.01, enabling us to reject the null hypothesis at the 1% level, affirming the relevance of the instruments. Moreover, we can also see in the first stage columns, the coefficients of instrumental variables are significant at the 1% level. Additionally, the Cragg-Donald Wald F-statistic significantly exceeds the threshold of 10 (Stock and Yogo, 2005), with our results showing a value well above this rule of thumb, which is commonly considered sufficient to rule out the presence of weak instruments. Lastly, the Hausman test gives p-value=0.015, which indicates the OLS estimate is affected by the endogeneity problem and is biased. Therefore, the 2SLS estimate is more reliable and should be used for inference first. This robust statistical grounding enhances the credibility of my model's estimations and their interpretations.

The results, in Table 1.6, indicate two angles that first, the OLS estimate is lower than it obtained through IV which suggests the presence of measurement error in the explanatory variable. This scenario is commonly referred to as attenuation bias, where inaccuracies in measuring the independent variable lead to an underestimation of its effect (Greene, 2000). Secondly, migrants in China confer a positive externality on native labour. Specifically, a 10 p.p. increase in the migrant share is associated with a 6.78% increase in the wage rate of local labour. This finding aligns with existing literature Altonji and Card (1991); Hanson and Slaughter (2002); Card and Lewis (2007); Lewis (2011); Xuguang and Jiajia (2022), which documents the positive effects of migrant labour on local wage levels. Following the assumption established by Combes et al. (2013), we assume that the agglomeration effects are identical for both migrant and native groups. So that if it is a positive coefficient for the migrant share in our analysis indicates that an increase in the proportion of migrants within a labour market is associated with higher wages for native workers. This relationship suggests that the presence of migrants positively affects the productivity of native workers, leading to increased wage levels. Such a dynamic is

Table 1.6: Comparing OLS Results and IV-Results

| VARIABLES | (1)<br>OLS _R_C<br>lnwage | (2)<br>IV_First _R_C<br>ms | (3)<br>IV_First _R_C<br>lnemp_den | (4)<br>IV_Second _R_C<br>lnwage |
|---|---|---|---|---|
| lmin_dis | | -0.013*** | 0.014*** | |
| | | (0.001) | (0.003) | |
| structure_90 | | 0.050*** | 0.128*** | |
| | | (0.003) | (0.018) | |
| ms | 0.446*** | | | 0.678*** |
| | (0.133) | | | (0.223) |
| lnemp_den | -0.049** | | | -0.046 |
| | (0.020) | | | (0.157) |
| educ | 0.065*** | -0.001** | 0.006** | 0.066*** |
| | (0.004) | (0.000) | (0.002) | (0.005) |
| male | 0.178*** | 0.002 | -0.024* | 0.177*** |
| | (0.013) | (0.002) | (0.013) | (0.013) |
| exp | 0.029*** | -0.001*** | 0.004* | 0.030*** |
| | (0.003) | (0.000) | (0.002) | (0.003) |
| exp_sq | -0.001*** | 0.000* | -0.000 | -0.001*** |
| | (0.000) | (0.000) | (0.000) | (0.000) |
| marriage | 0.067*** | 0.013*** | -0.006 | 0.060*** |
| | (0.022) | (0.004) | (0.024) | (0.018) |
| Constant | 3.223*** | -0.334*** | 0.673** | 3.226*** |
| | (0.428) | (0.040) | (0.261) | (0.396) |
| City Control | YES | YES | YES | YES |
| Occupation Control | YES | YES | YES | YES |
| $N$ | 9,086 | 9,061 | 9,061 | 9,061 |
| $R^2$ | 0.266 | | | 0.264 |
| IV F-stat | | | | 157.19 |
| Anderson LM | | | | 80.637 |
| CD Wald F-statistic | | | | 33.05 |
| Hausman Test | | | | 0.015 |

Note: The first-stage results of 2SLS are in Columns (2) and (3). Columns (1) and (4) are the OLS and second-stage results with robust-cluster errors. Robust standard errors in parentheses: ***Significant at 1%; **Significant at 5%; *Significant at 10%. The full table result is in Table 1.12.

typically interpreted as complementarity between migrants and natives, implying that the two groups synergistically enhance each other's productivity, likely due to their different yet complementary skills within the labour market. Conversely, a negative coefficient would imply substitutability between migrants and native workers. In such scenarios, an increase in the migrant share could be associated with lower wages for native workers. This would suggest that migrants are perceived as direct competitors, filling roles that natives could occupy, which could exert downward pressure on wages due to increased labour supply. This effect underscores the importance of considering the skill overlap between migrant and native workers when evaluating the impact of immigration on local labour markets. our analysis reveals a complementarity situation between domestic immigrants and local labour. Our study is corroborated by the findings of Peri (2007); Peri and Sparber (2009);

Ortega and Verdugo (2011); Ottaviano and Peri (2012); Combes et al. (2015), and provides robust evidence supporting the hypothesis that migrant labour not only contributes to the wage growth of native labour through direct labour market interactions but also through enhanced agglomeration effects that benefit the broader economic environment.

The relationship between the proportion of immigrants and local wage levels is not linear. Still, it displays a nonlinear, U-shaped pattern, as demonstrated by (Ioramashvili, 2023). Initially, an influx of immigrants might lead to increased inequality and a negative impact on local wages. This early-stage effect can be attributed to the adjustment costs and crowded effect associated with integrating new workers into the labour market, where an increase in labour supply may temporarily depress wages (Burstein et al., 2020). However, over time, as immigrant communities establish stronger networks (an increase in the migrant share) and the local labour market adapts to the increased labour pool, the impact of immigration shifts (Du et al., 2005; McKenzie and Rapoport, 2007). In the longer term, immigration tends to contribute positively to local wages, particularly in urban settings where the dynamics of labour demand and the diversity of job opportunities can more readily absorb and benefit from the skills and contributions of immigrants (Ioramashvili, 2023). This phase reflects a reduction in inequality as the labour market becomes more inclusive and as immigrants and natives alike benefit from the resultant economic growth and labour market efficiencies.

In this section, our objective is to employ OLS regression, incorporating a quadratic term for migrant share to investigate the presence of a nonlinear relationship between the proportion of immigrants and local wage levels. After incorporating the quadratic term, we further include the quadratic terms of the instrumental variables, presented in Table 1.6, and a historical cities dummy variable as instrumental variables in the two-stage regression model. In Table 1.7, column (5), the instrumental variables satisfy all model specifications, with the number of instrumental variables exceeding that of endogenous variables. Importantly, the Hansen p-value of 0.921 indicates that the model does not violate the null hypothesis, suggesting a good fit and validity of the instruments used. By comparing the results in columns (1) and (5), we can see that the coefficient of the quadratic form of migrant share is significant at 10% level after being instrumented.

The results align with established assumptions that permit the quadratic equation to exhibit a non-significant first-order term as outlined in Haans et al. (2016). We acknowledge that in our quadratic regression analyses, both OLS and TSLS, the coefficient of the linear term is not statistically significant, indicating that its effect cannot be reliably distinguished from zero given the current data and model

specification. This lack of statistical significance calls for caution when interpreting the point estimate of the linear term, as it may not accurately reflect the underlying relationship. Conversely, the quadratic term remains statistically significant and consistent with theoretical expectations, reinforcing its importance in capturing the higher-order effects within the model. For the current analysis, the negligible role of the linear term suggests a focus on the quadratic term to understand the primary dynamics of the relationship. This U-shaped pattern suggests that initial increases in the migrant share may decrease the local hourly wage rate up to a certain point, after which further increases in the migrant share lead to increases in the native hourly wage rate.

When employing OLS estimation, we identify a turning point at 0.073. In practical terms, this suggests that once the migrant share reaches 7.3%, the marginal impact of immigration on the wages of native labour transitions from negative to positive. This finding provides an initial estimation of the threshold at which the influence of immigration begins to benefit native workers in terms of wage effects. To address potential endogeneity issues, we implement an Instrumental Variables (IV) estimation, as detailed in Section 1.4.1. This approach allows us to refine our understanding of the relationship and determine a more accurate turning point. After applying IV estimation, the turning point shifts to 0.2732, indicating that the actual threshold for a positive impact on native wages occurs when the migrant share reaches 27.32%. This substantial difference in turning points between the OLS and IV estimations highlights the importance of addressing endogeneity to uncover the true effects of immigration. The quadratic term estimations for both OLS and IV are visually represented in the graphs shown in Figures 1.1 and 1.2. These graphs illustrate the U-shaped relationships and the respective turning points, providing a clear visual representation of how the impact of immigration on native wages varies with changes in the proportion of immigrants.

This observed pattern demonstrates a nonlinear association between the migration share and the natural logarithm of the local wage rate, characterised by an initial decrease followed by a subsequent increase in the marginal effect. This non-monotonic nature of the relationship between these variables underscores a transformation in the dynamic between local and migrant labour, evolving from a simplistic view of mutual exclusivity or complementarity to a more structurally integrated coexistence.

The impact of migration on local wage levels follows a phased pattern, reflecting a process of continuous adaptation within urban labour markets. At lower levels of migrant share, specifically, in our case, when the average proportion of urban immigrants is below 27.32% in China, the influx of migrants tends to negatively impact the wages of native workers. This observation aligns with previous findings in

the literature that highlight increased competition for jobs between immigrants and the local labor force (Knight and Yueh, 2009; Burstein et al., 2020). Initially, immigration expands the local labour supply, capital inflows and technological advancements may not adjust swiftly enough, exerting downward pressure on wages. However, as the continuous influx of migrants grows incrementally over time and

Table 1.7: Quadratic form results of OLS and IV Estimation

| VARIABLES | (1) OLS _R_C lnwage | (2) IV_First _R_C ms | (3) IV_First _R_C ms_sq | (4) IV_First _R_C lnemp_den | (5) IV_Second _R_C lnwage |
|---|---|---|---|---|---|
| lmin_dis | | -0.032*** | -0.096*** | -0.164*** | |
| | | (0.002) | (0.013) | (0.014) | |
| historic | | 0.051*** | -0.004 | -0.034** | |
| | | (0.004) | (0.032) | (0.015) | |
| structure_90 | | -0.020** | 0.145* | 0.723*** | |
| | | (0.010) | (0.077) | (0.091) | |
| lmin_dis_sq | | -0.003*** | 0.005** | -0.030*** | |
| | | (0.000) | (0.002) | (0.002) | |
| structure_90_sq | | -0.012*** | -0.237*** | -0.262*** | |
| | | (0.003) | (0.026) | (0.027) | |
| ms | -0.135 | | | | -0.036 |
| | (0.070) | | | | (0.299) |
| ms_sq | 0.925** | | | | 0.066* |
| | (0.111) | | | | (0.034) |
| lnemp_den | -0.051** | | | | -0.055 |
| | (0.012) | | | | (0.047) |
| educ | 0.065*** | -0.001*** | -0.005 | 0.006** | 0.065*** |
| | (0.006) | (0.000) | (0.003) | (0.002) | (0.003) |
| male | 0.177*** | 0.003 | 0.022 | -0.020 | 0.177*** |
| | (0.014) | (0.002) | (0.016) | (0.013) | (0.013) |
| exp | 0.029*** | -0.001*** | -0.005* | 0.004* | 0.030*** |
| | (0.002) | (0.000) | (0.003) | (0.002) | (0.003) |
| exp_sq | -0.001** | 0.000* | 0.000 | -0.000* | -0.001*** |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| marriage | 0.065 | 0.013*** | 0.098*** | -0.022 | 0.063*** |
| | (0.028) | (0.004) | (0.028) | (0.024) | (0.024) |
| Constant | 3.310*** | -0.301*** | -5.358*** | -0.321 | 3.341*** |
| | (0.323) | (0.043) | (0.379) | (0.278) | (0.244) |
| City Control | YES | YES | YES | YES | YES |
| Occupation Control | YES | YES | YES | YES | YES |
| N | 9,086 | 9,061 | 9,061 | 9,061 | 9,061 |
| $R^2$ | 0.268 | | | | 0.267 |
| IV F-stat | | | | | 149.7 |
| Anderson LM | | | | | 321.3 |
| CD Wald F-statistic | | | | | 91.36 |
| Hansen p-value | | | | | 0.921 |
| Hausman Test | | | | | 0.051 |

Note: The first-stage results of IV are in columns (2), (3), and (4), column (5) is the second-stage result. In the brackets are the robust-clustered standard errors. Robust standard errors in parentheses: ***Significant at 1%; **Significant at 5%; *Significant at 10%. The full table result is in Tables 1.15 and 1.16.

Figure 1.1: OLS Quadratic Estimation
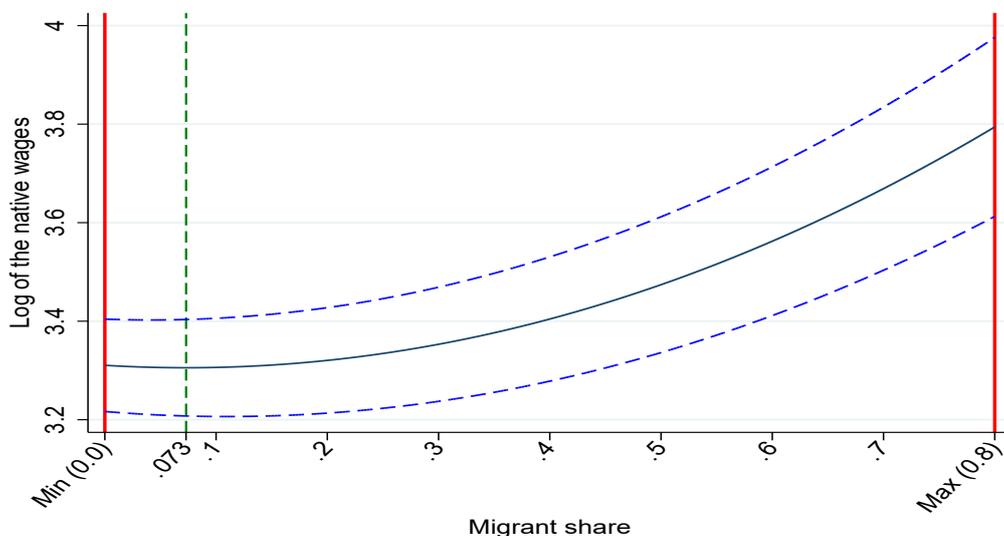


Note: This figure illustrates the quadratic relationship between native wages (logarithmic scale) and migrant share. The solid curve line shows the estimated trend, with red lines marking the minimum (0.0) and maximum (0.8) migrant share values. The green dashed line highlights a critical value (turning point) of 0.073. The blue dashed lines represent the upper and lower bounds of the 90% confidence intervals.

Figure 1.2: IV Quadratic Estimation
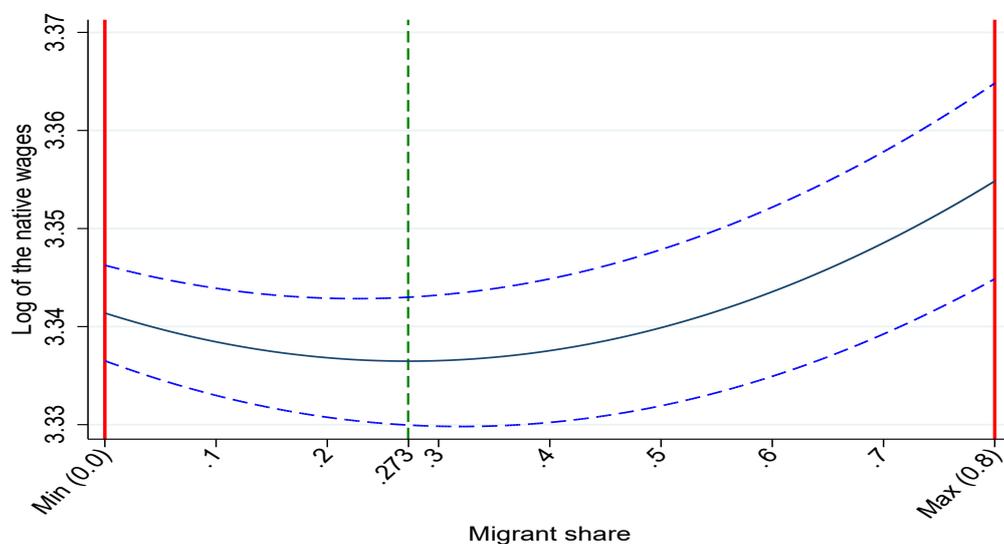


Note: This figure illustrates the quadratic relationship between native wages (logarithmic scale) and migrant share. The solid curve line shows the estimated trend, with red lines marking the minimum (0.0) and maximum (0.8) migrant share values. The green dashed line highlights a critical value (turning point) of 0.273. The blue dashed lines represent the upper and lower bounds of the 90% confidence intervals.

certain cities reach advanced levels of urbanisation, a balance between industrial demand and labour supply begins to emerge. Our findings indicate that in China, once the migrant share surpasses 27.32%, the labour market dynamics adjust, leading to an increase in local wage rates. This equilibrium typically arises when

urban development reaches a stage where the economic infrastructure and market dynamics are mature enough to effectively absorb and utilise the available workforce, consistent with insights from Du et al. (2005); McKenzie and Rapoport (2007), and benefits both local workers and immigrants, as noted by (Ioramashvili, 2023). During this adaptive phase, migrants engage in productive competition with the local workforce, and in specific sectors such as caregiving and labour-intensive industries, they assume a complementary role. This development supports findings from studies like those by Peri (2007); Peri and Sparber (2009); Ortega and Verdugo (2011), which illustrate the complex and beneficial interactions between migrant and native labour forces that significantly contribute to the economic vitality of urban centres.

## 1.6 Robustness Check

In this section, we explore how variations in the educational levels of immigrant labour affect the hourly wages of local labour. Theoretically, local workers who are more easily substitutable by immigrant labour within the production function are likely to experience less positive, or even negative, impacts from immigration. Drawing from the existing literature, such as Hall and Farkas (2008), we categorise migrant workers with 12 years of education or less as low-skilled, and those with more than 12 years as high-skilled. This classification enables a comprehensive analysis of how differing levels of immigrant skill influence local labour markets, potentially exacerbating competition for certain job types or enhancing overall productivity through complementarity in higher-skill domains.

The influence of migrant share on the wages of high-skilled and low-skilled native workers is detailed in two separate tables. Table 1.8 presents the results for high-skilled native wage impacts, while Table 1.9 presents on the impacts to low-skilled native wages. Within each table, columns (2) and (3) display the first-step estimation results for each group separately. Columns (1) and (4) provide the results from the OLS and IV second-step estimations conducted separately on skilled and unskilled native workers. In each IV estimation, the model specifications meet the critical values, indicating that the instrumental variables demonstrate high relevance in the first stage and successfully pass the weak-identification tests. This outcome ensures that the instruments used are appropriately influencing the endogenous variables, thereby validating the strength and reliability of the instrumental variable approach within our analysis.

Following expectations, our findings indicate that the impact of migrants on wages is significant and positive exclusively for skilled natives. Although the influence on unskilled workers is not statistically significant, attributable to a lack of precision in the estimates, the OLS estimation still reveals a positive correlation

Table 1.8: Migrant Share's Influence to High Skilled Native Wage

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | OLS_R_C High_skilled | IV_First _R_C High_skilled | IV_First _R_C High_skilled | IV_Second _R_C High_skilled |
| VARIABLES | lnwage | ms | lnemp_den | lnwage |
| lmin_dis | | -0.016*** | 0.008 | |
| | | (0.001) | (0.005) | |
| structure_90 | | 0.065*** | 0.164*** | |
| | | (0.004) | (0.025) | |
| ms | 0.558*** | | | 0.472* |
| | (0.090) | | | (0.264) |
| lnemp_den | -0.073*** | | | -0.017 |
| | (0.016) | | | (0.165) |
| educ | 0.182*** | 0.002 | 0.009 | 0.181*** |
| | (0.009) | (0.002) | (0.009) | (0.009) |
| male | 0.123*** | 0.006* | -0.009 | 0.122*** |
| | (0.019) | (0.003) | (0.021) | (0.020) |
| exp | 0.040*** | -0.000 | 0.007 | 0.039*** |
| | (0.004) | (0.001) | (0.004) | (0.004) |
| exp_sq | -0.001*** | -0.000 | -0.000 | -0.001*** |
| | (0.000) | (0.000) | (0.000) | (0.000) |
| marriage | 0.017 | 0.009* | -0.029 | 0.017 |
| | (0.032) | (0.006) | (0.034) | (0.033) |
| Constant | 0.769* | -0.465*** | -0.373 | 0.750* |
| | (0.397) | (0.073) | (0.445) | (0.403) |
| N | 3,433 | 3,426 | 3,426 | 3,426 |
| $R^2$ | 0.361 | | | 0.358 |
| IV F-stat | | | | 103.3 |
| Anderson LM | | | | 31.77 |
| CD Wald F-statistic | | | | 15.95 |

Note: The first-stage results of 2SLS are in Columns (2) and (3). Columns (1) and (4) are the OLS and second-stage results with robust-cluster errors. Robust standard errors in parentheses: ***Significant at 1%; **Significant at 5%; *Significant at 10%. The full table is in Table 1.13.

Table 1.9: Migrant Share's Influence to Low Skilled Native Wage

| VARIABLES | (1) OLS_R_C Low_skilled lnwage | (2) IV_First _R_C Low_skilled ms | (3) IV_First _R_C Low_skilled lnemp_den | (4) IV_Second _R_C Low_skilled lnwage |
|---|---|---|---|---|
| lmin_dis | | -0.010*** | 0.020*** | |
| | | (0.001) | (0.004) | |
| structure_90 | | 0.035*** | 0.089*** | |
| | | (0.003) | (0.025) | |
| ms | 0.220*** | | | 0.503 |
| | (0.081) | | | (0.370) |
| lnemp_den | -0.041*** | | | 0.047 |
| | (0.012) | | | (0.170) |
| educ | 0.021*** | -0.002*** | 0.005 | 0.021*** |
| | (0.004) | (0.001) | (0.004) | (0.004) |
| male | 0.253*** | -0.001 | -0.031* | 0.257*** |
| | (0.017) | (0.003) | (0.018) | (0.018) |
| exp | 0.027*** | -0.001* | 0.002 | 0.027*** |
| | (0.003) | (0.001) | (0.004) | (0.004) |
| exp_sq | -0.001*** | 0.000 | -0.000 | -0.001*** |
| | (0.000) | (0.000) | (0.000) | (0.000) |
| marriage | 0.061* | 0.012** | 0.003 | 0.054 |
| | (0.033) | (0.005) | (0.036) | (0.033) |
| Constant | 4.484*** | -0.192*** | 1.415*** | 4.340*** |
| | (0.296) | (0.051) | (0.326) | (0.412) |
| $N$ | 5,653 | 5,635 | 5,635 | 5,635 |
| $R^2$ | 0.112 | | | 0.101 |
| IV F-stat | | | | 38.03 |
| Anderson LM | | | | 29.47 |
| CD F-statistic | | | | 14.76 |

Note: The first-stage results of 2SLS are in Columns (2) and (3). Columns (1) and (4) are the OLS and second-stage results with robust-cluster errors. Robust standard errors in parentheses: ***Significant at 1%; **Significant at 5%; *Significant at 10%. The full table is in Table 1.14.

between the migrant share and the wage rate of low-skilled workers. Nevertheless, the primary impact of the migrant share is quantified at 0.472 for skilled natives, suggesting that, on average, a 10 p.p. increase in the migrant share results in a 4.72% wage increase for local skilled labour. These outcomes resonate with the premise that skilled natives, being less substitutable by migrants than unskilled natives, derive greater benefits from the presence of migrants (Combes et al., 2015). This scenario supports the concept that internal migration typically boosts the wages of urban native workers, with the effects being more pronounced for high-skilled native workers (Combes et al., 2020; Zhao, 2020). This enhanced impact for skilled workers can be attributed to complementarities in production functions and the advantages stemming from agglomeration economies, which collectively facilitate a more substantial wage uplift for skilled natives within urban labour markets.

## 1.7 Policy Implications

The analysis highlights the intricate relationship between migration and its impact on local wages in urban China, emphasising the need for evidence-based policies to maximise migration benefits while addressing challenges, particularly for unskilled workers.

The authorities should enhance labour market integration for migrants. Implement targeted vocational training and education programs for unskilled workers to improve their employability and adaptability to evolving labour market conditions. Prioritise certification of prior skills and sector-specific upskilling initiatives to integrate migrants into productive economic roles effectively. Further, they should address short-term disparities for unskilled workers. Introduce short-term financial support mechanisms, such as wage subsidies or stipends, aimed at unskilled workers during labour market adjustments. These measures can mitigate negative short-term impacts and provide a foundation for sustained economic participation.

The government should navigate and promote balanced migrant distribution. High concentrations of migrants in specific urban areas can saturate labour markets, reducing wage benefits and exacerbating inequalities. To address this, implement measures to encourage settlement in underdeveloped regions or smaller cities. Incentives for relocation, combined with support infrastructures such as affordable housing and accessible social services, can make these areas more appealing to migrants. Collaborative efforts with local governments will be critical in achieving balanced geographical distribution and fostering inclusive economic growth.

## 1.8 Conclusion

This study provides a detailed analysis of the impact of internal migration on the wages of native workers in Chinese cities, utilising both linear and nonlinear models, alongside an IV approach to address potential endogeneity. In comparison to existing literature, which often focuses on the impacts of international migration on wage dynamics, this study contributes by examining the unique context of internal migration within China. The findings corroborate previous Chinese internal studies, such as by Combes et al. (2015); Xuguang and Jiajia (2022), which highlight the complexity of migration's effects on local labour markets. However, this study goes further by offering insights specific to the Chinese urban context, particularly the turning point at which migration shifts from having a negative to a positive effect on wages, a threshold rarely addressed in the literature.

The linear model reveals a significant positive impact, showing that a 10 p.p. increase in the migrant share leads to a 6.78% rise in local wages, reflecting the competitive pressures and labour dynamics within urban China. However, unlike many prior studies that present a linear relationship, this research uncovers a nonlinear dynamic, when the migrant share exceeds 27.32%, the initially negative effect on wages reverses, resulting in wage increases, particularly benefiting skilled workers. This nonlinearity suggests that migration has context-dependent effects, particularly in transitioning economies like China. Skilled workers, in particular, gain more from migration, as a 10 p.p. increase in migrant share leads to a 4.72% wage increase for them. In contrast, the effect on unskilled workers, while still positive, is less substantial and statistically less certain, echoing findings by Combes et al. (2015).

This study's novel contribution lies in its methodological approach. By employing both linear and nonlinear models and IV strategy to account for endogeneity, the study not only verifies the robustness of its results but also introduces a more precise understanding of the threshold where migration's effects become beneficial. The identification strategy, utilising city-level instruments such as geographical proximity to seaports and historical population data, provides a clearer causal interpretation of the results, addressing concerns about reverse causality between wages and migration.

From a policy perspective, these findings highlight the need for targeted interventions to maximise the positive effects of migration while mitigating its short-term negative impacts, particularly for unskilled workers. Policies focused on integrating migrants into the labour market through education, vocational training, and support services will be crucial. Additionally, promoting a more even

distribution of migrants across urban centres can prevent labour market saturation and ensure more equitable benefits from migration. Special attention should be given to unskilled workers who may be more vulnerable to the early-stage negative impacts of migration.

While this study provides valuable insights, several limitations must be acknowledged. The cross-sectional nature of the data limits the ability to track how migration's effects on wages evolve over time. Additionally, potential measurement errors in key variables like wage levels and migrant shares could influence the accuracy of the findings. Although instrumental variables were employed to address endogeneity, concerns about the strength and validity of these instruments remain. Violations of the exclusion restriction assumption could lead to biased estimates, highlighting the need for further validation through longitudinal data and more refined identification techniques.

In conclusion, this study contributes significantly to the literature by offering a more complex understanding of internal migration in China. It supports the argument that, under appropriate management, migration can be a driver of economic growth and wage improvement in urban areas. The insights into nonlinear wage impacts and the nuanced effects on different skill levels provide a strong foundation for future research and policy interventions. Moreover, the identification strategy enriches the methodological toolbox for migration studies.

# 1.A  Appendix

## 1.A.1  Additional Tables

Table 1.10: Full Table OLS Estimation Results

| VARIABLES | (1) lnwage | (2) lnwage | (3) lnwage | (4) lnwage | (5) lnwage |
|---|---|---|---|---|---|
| ms | 0.397*** | 0.423*** | 0.427*** | 0.456*** | 0.446*** |
|  | (0.066) | (0.063) | (0.063) | (0.062) | (0.062) |
| lnemp_den | -0.041*** | -0.048*** | -0.045*** | -0.049*** | -0.049*** |
|  | (0.011) | (0.010) | (0.010) | (0.010) | (0.010) |
| educ |  | 0.052*** | 0.052*** | 0.066*** | 0.065*** |
|  |  | (0.002) | (0.002) | (0.002) | (0.002) |
| male |  |  | 0.179*** | 0.176*** | 0.178*** |
|  |  |  | (0.013) | (0.013) | (0.013) |
| exp |  |  |  | 0.033*** | 0.029*** |
|  |  |  |  | (0.002) | (0.002) |
| exp_sq |  |  |  | -0.001*** | -0.001*** |
|  |  |  |  | (0.000) | (0.000) |
| marriage |  |  |  |  | 0.067*** |
|  |  |  |  |  | (0.023) |
| lnpubexp | 0.111*** | 0.035 | 0.034 | -0.020 | -0.018 |
|  | (0.038) | (0.037) | (0.036) | (0.036) | (0.036) |
| lnexpst | 0.070*** | 0.067*** | 0.072*** | 0.077*** | 0.078*** |
|  | (0.012) | (0.011) | (0.011) | (0.011) | (0.011) |
| lnexpedu | 0.005 | 0.056 | 0.047 | 0.078** | 0.077** |
|  | (0.036) | (0.034) | (0.034) | (0.033) | (0.033) |
| lnfixedasset | -0.057*** | -0.055*** | -0.053*** | -0.052*** | -0.054*** |
|  | (0.018) | (0.017) | (0.017) | (0.017) | (0.017) |
| Occ1 | 0.508*** | 0.298*** | 0.287*** | 0.203*** | 0.203*** |
|  | (0.040) | (0.040) | (0.039) | (0.039) | (0.039) |
| Occ2 | 0.417*** | 0.260*** | 0.255*** | 0.234*** | 0.234*** |
|  | (0.025) | (0.025) | (0.025) | (0.024) | (0.024) |
| Occ3 | 0.213*** | 0.075*** | 0.095*** | 0.074*** | 0.074*** |
|  | (0.025) | (0.025) | (0.024) | (0.024) | (0.024) |
| Occ4 | -0.064** | -0.088*** | -0.058** | -0.055** | -0.053** |
|  | (0.025) | (0.025) | (0.024) | (0.024) | (0.024) |
| Occ5 | -0.054 | 0.034 | 0.044 | 0.044 | 0.043 |
|  | (0.059) | (0.057) | (0.056) | (0.055) | (0.055) |
| Occ6 | 0.011 | 0.041* | 0.048** | 0.038 | 0.038 |
|  | (0.026) | (0.025) | (0.024) | (0.024) | (0.024) |
| Occ7 | 0.409** | 0.156 | 0.092 | 0.190 | 0.192 |
|  | (0.200) | (0.193) | (0.191) | (0.187) | (0.187) |
| Constant | 3.460*** | 3.465*** | 3.358*** | 3.217*** | 3.223*** |
|  | (0.255) | (0.246) | (0.243) | (0.239) | (0.239) |
| $N$ | 9,086 | 9,086 | 9,086 | 9,086 | 9,086 |
| $R^2$ | 0.159 | 0.219 | 0.235 | 0.265 | 0.266 |

Note: ***Significant at 1%; **Significant at 5%; *Significant at 10%.

Table 1.11: Ordinary and Adjusted OLS Estimation Results

| VARIABLES | (1) OLS lnwage | (2) OLS_R lnwage | (3) OLS_R_C lnwage |
|---|---|---|---|
| ms | 0.446*** | 0.446*** | 0.446*** |
| | (0.062) | (0.064) | (0.133) |
| lnemp_den | -0.049*** | -0.049*** | -0.049** |
| | (0.010) | (0.010) | (0.020) |
| educ | 0.065*** | 0.065*** | 0.065*** |
| | (0.002) | (0.003) | (0.004) |
| male | 0.178*** | 0.178*** | 0.178*** |
| | (0.013) | (0.013) | (0.013) |
| exp | 0.029*** | 0.029*** | 0.029*** |
| | (0.002) | (0.003) | (0.003) |
| exp_sq | -0.001*** | -0.001*** | -0.001*** |
| | (0.000) | (0.000) | (0.000) |
| marriage | 0.067*** | 0.067*** | 0.067*** |
| | (0.023) | (0.024) | (0.022) |
| lnpubexp | -0.018 | -0.018 | -0.018 |
| | (0.036) | (0.036) | (0.071) |
| lnexpst | 0.078*** | 0.078*** | 0.078*** |
| | (0.011) | (0.011) | (0.020) |
| lnexpedu | 0.077** | 0.077** | 0.077 |
| | (0.033) | (0.033) | (0.063) |
| lnfixedasset | -0.054*** | -0.054*** | -0.054 |
| | (0.017) | (0.017) | (0.035) |
| Occ1 | 0.203*** | 0.203*** | 0.203*** |
| | (0.039) | (0.042) | (0.049) |
| Occ2 | 0.234*** | 0.234*** | 0.234*** |
| | (0.024) | (0.026) | (0.028) |
| Occ3 | 0.074*** | 0.074*** | 0.074*** |
| | (0.024) | (0.025) | (0.027) |
| Occ4 | -0.053** | -0.053** | -0.053* |
| | (0.024) | (0.025) | (0.028) |
| Occ5 | 0.043 | 0.043 | 0.043 |
| | (0.055) | (0.063) | (0.066) |
| Occ6 | 0.038 | 0.038 | 0.038 |
| | (0.024) | (0.025) | (0.028) |
| Occ7 | 0.192 | 0.192 | 0.192 |
| | (0.187) | (0.208) | (0.231) |
| Constant | 3.223*** | 3.223*** | 3.223*** |
| | (0.239) | (0.236) | (0.428) |
| $N$ | 9,086 | 9,086 | 9,086 |
| $R^2$ | 0.266 | 0.266 | 0.266 |

Note: ***Significant at 1%; **Significant at 5%; *Significant at 10%.

Table 1.12: Comparing OLS Results and Second Stage IV-Results

| VARIABLES | (1)<br>OLS _R_C<br>lnwage | (2)<br>IV_First _R_C<br>ms | (3)<br>IV_First _R_C<br>lnemp_den | (4)<br>IV_Second _R_C<br>lnwage |
|---|---|---|---|---|
| lmin_dis | | -0.013*** | 0.014*** | |
| | | (0.001) | (0.003) | |
| structure_90 | | 0.050*** | 0.128*** | |
| | | (0.003) | (0.018) | |
| ms | 0.446*** | | | 0.678*** |
| | (0.133) | | | (0.223) |
| lnemp_den | -0.049** | | | -0.046 |
| | (0.020) | | | (0.157) |
| educ | 0.065*** | -0.001** | 0.006** | 0.066*** |
| | (0.004) | (0.000) | (0.002) | (0.005) |
| male | 0.178*** | 0.002 | -0.024* | 0.177*** |
| | (0.013) | (0.002) | (0.013) | (0.013) |
| exp | 0.029*** | -0.001*** | 0.004* | 0.030*** |
| | (0.003) | (0.000) | (0.002) | (0.003) |
| exp_sq | -0.001*** | 0.000* | -0.000 | -0.001*** |
| | (0.000) | (0.000) | (0.000) | (0.000) |
| marriage | 0.067*** | 0.013*** | -0.006 | 0.060*** |
| | (0.022) | (0.004) | (0.024) | (0.018) |
| lnpubexp | -0.018 | 0.108*** | 0.378*** | -0.041 |
| | (0.071) | (0.006) | (0.038) | (0.089) |
| lnexpst | 0.078*** | 0.078*** | -0.000 | 0.061*** |
| | (0.020) | (0.002) | (0.010) | (0.020) |
| lnexpedu | 0.077 | 0.038*** | -0.267*** | 0.080 |
| | (0.063) | (0.006) | (0.043) | (0.087) |
| lnfixedasset | -0.054 | -0.153*** | 0.244*** | -0.027 |
| | (0.035) | (0.003) | (0.020) | (0.067) |
| Occ1 | 0.203*** | -0.001 | -0.014 | 0.205*** |
| | (0.049) | (0.006) | (0.044) | (0.047) |
| Occ2 | 0.234*** | -0.004 | -0.068** | 0.234*** |
| | (0.028) | (0.004) | (0.027) | (0.034) |
| Occ3 | 0.074*** | 0.005 | -0.056** | 0.074** |
| | (0.027) | (0.004) | (0.027) | (0.034) |
| Occ4 | -0.053* | -0.008** | -0.095*** | -0.050 |
| | (0.028) | (0.004) | (0.028) | (0.034) |
| Occ5 | 0.043 | -0.007 | -0.116** | 0.047 |
| | (0.066) | (0.009) | (0.051) | (0.075) |
| Occ6 | 0.038 | -0.001 | -0.054** | 0.038 |
| | (0.028) | (0.004) | (0.027) | (0.029) |
| Occ7 | 0.192 | 0.008 | -0.211 | 0.190 |
| | (0.231) | (0.029) | (0.176) | (0.230) |
| Constant | 3.223*** | -0.334*** | 0.673** | 3.226*** |
| | (0.428) | (0.040) | (0.261) | (0.396) |
| $N$ | 9,086 | 9,061 | 9,061 | 9,061 |
| $R^2$ | 0.266 | | | 0.264 |
| IV F-stat | | | | 157.19 |
| Anderson LM | | | | 80.637 |
| CD Wald F-statistic | | | | 33.05 |

Note: The first-stage results of 2SLS are in Columns (2) and (3). Columns (1) and (4) are the OLS and second-stage results with robust-cluster errors. Robust standard errors in parentheses: ***Significant at 1%; **Significant at 5%; *Significant at 10%.

Table 1.13: High Skilled Migrant Worker's Influence to Native Wage

| VARIABLES | (1)<br>OLS_R_C<br>High_skilled<br>lnwage | (2)<br>IV_First _R_C<br>High_skilled<br>ms | (3)<br>IV_First _R_C<br>High_skilled<br>lnemp_den | (4)<br>IV_Second _R_C<br>High_skilled<br>lnwage |
|---|---|---|---|---|
| lmin_dis | | -0.016*** | 0.008 | |
| | | (0.001) | (0.005) | |
| structure_90 | | 0.065*** | 0.164*** | |
| | | (0.004) | (0.025) | |
| ms | 0.558*** | | | 0.472* |
| | (0.090) | | | (0.264) |
| lnemp_den | -0.073*** | | | -0.017 |
| | (0.016) | | | (0.165) |
| educ | 0.182*** | 0.002 | 0.009 | 0.181*** |
| | (0.009) | (0.002) | (0.009) | (0.009) |
| male | 0.123*** | 0.006* | -0.009 | 0.122*** |
| | (0.019) | (0.003) | (0.021) | (0.020) |
| exp | 0.040*** | -0.000 | 0.007 | 0.039*** |
| | (0.004) | (0.001) | (0.004) | (0.004) |
| exp_sq | -0.001*** | -0.000 | -0.000 | -0.001*** |
| | (0.000) | (0.000) | (0.000) | (0.000) |
| marriage | 0.017 | 0.009* | -0.029 | 0.017 |
| | (0.032) | (0.006) | (0.034) | (0.033) |
| lnpubexp | 0.065 | 0.120*** | 0.623*** | 0.045 |
| | (0.056) | (0.010) | (0.061) | (0.093) |
| lnexpst | 0.068*** | 0.086*** | -0.019 | 0.074*** |
| | (0.018) | (0.003) | (0.018) | (0.028) |
| lnexpedu | 0.035 | 0.039*** | -0.460*** | 0.060 |
| | (0.054) | (0.010) | (0.060) | (0.088) |
| lnfixedasset | -0.039 | -0.166*** | 0.242*** | -0.065 |
| | (0.026) | (0.004) | (0.025) | (0.070) |
| Occ1 | 0.158*** | 0.002 | 0.122** | 0.147** |
| | (0.055) | (0.010) | (0.059) | (0.059) |
| Occ2 | 0.192*** | 0.003 | 0.029 | 0.186*** |
| | (0.044) | (0.008) | (0.047) | (0.044) |
| Occ3 | 0.042 | 0.009 | 0.034 | 0.036 |
| | (0.044) | (0.008) | (0.047) | (0.044) |
| Occ4 | -0.062 | 0.001 | 0.031 | -0.067 |
| | (0.048) | (0.008) | (0.052) | (0.049) |
| Occ5 | -0.040 | -0.009 | 0.060 | -0.047 |
| | (0.172) | (0.030) | (0.183) | (0.172) |
| Occ6 | -0.069 | 0.013 | 0.050 | -0.074 |
| | (0.053) | (0.009) | (0.057) | (0.054) |
| Occ7 | 0.256 | 0.036 | 0.083 | 0.249 |
| | (0.200) | (0.035) | (0.213) | (0.201) |
| Constant | 0.769* | -0.465*** | -0.373 | 0.750* |
| | (0.397) | (0.073) | (0.445) | (0.403) |
| N | 3,433 | 3,426 | 3,426 | 3,426 |
| $R^2$ | 0.361 | | | 0.358 |
| IV F-stat | | | | 103.3 |
| Anderson LM | | | | 31.77 |
| CD Wald F-statistic | | | | 15.95 |

Note: The first-stage results are in Columns (2) and (3). Columns (1) and (4) are the OLS and second-stage results. Robust standard errors in parentheses: ***1%, **5%, *10%.

Table 1.14: Low Skilled Migrant Worker's Influence to Native Wage

| VARIABLES | (1)<br>OLS_R_C<br>Low_skilled<br>lnwage | (2)<br>IV_First _R_C<br>Low_skilled<br>ms | (3)<br>IV_First _R_C<br>Low_skilled<br>lnemp_den | (4)<br>IV_Second _R_C<br>Low_skilled<br>lnwage |
|---|---|---|---|---|
| lmin_dis | | -0.010*** | 0.020*** | |
| | | (0.001) | (0.004) | |
| structure_90 | | 0.035*** | 0.089*** | |
| | | (0.003) | (0.025) | |
| ms | 0.220*** | | | 0.503 |
| | (0.081) | | | (0.370) |
| lnemp_den | -0.041*** | | | 0.047 |
| | (0.012) | | | (0.170) |
| educ | 0.021*** | -0.002*** | 0.005 | 0.021*** |
| | (0.004) | (0.001) | (0.004) | (0.004) |
| male | 0.253*** | -0.001 | -0.031* | 0.257*** |
| | (0.017) | (0.003) | (0.018) | (0.018) |
| exp | 0.027*** | -0.001* | 0.002 | 0.027*** |
| | (0.003) | (0.001) | (0.004) | (0.004) |
| exp_sq | -0.001*** | 0.000 | -0.000 | -0.001*** |
| | (0.000) | (0.000) | (0.000) | (0.000) |
| marriage | 0.061* | 0.012** | 0.003 | 0.054 |
| | (0.033) | (0.005) | (0.036) | (0.033) |
| lnpubexp | -0.121*** | 0.085*** | 0.196*** | -0.155*** |
| | (0.045) | (0.008) | (0.052) | (0.057) |
| lnexpst | 0.094*** | 0.074*** | 0.010 | 0.073** |
| | (0.013) | (0.002) | (0.012) | (0.030) |
| lnexpedu | 0.086** | 0.036*** | -0.161*** | 0.096** |
| | (0.041) | (0.007) | (0.052) | (0.044) |
| lnfixedasset | -0.031 | -0.136*** | 0.275*** | -0.023 |
| | (0.022) | (0.004) | (0.028) | (0.056) |
| Occ1 | 0.130** | 0.002 | -0.151** | 0.146** |
| | (0.064) | (0.010) | (0.076) | (0.070) |
| Occ2 | 0.170*** | -0.007 | -0.091*** | 0.180*** |
| | (0.030) | (0.005) | (0.034) | (0.034) |
| Occ3 | 0.099*** | 0.007 | -0.074** | 0.104*** |
| | (0.029) | (0.005) | (0.033) | (0.032) |
| Occ4 | -0.016 | -0.009** | -0.127*** | -0.001 |
| | (0.027) | (0.004) | (0.032) | (0.036) |
| Occ5 | 0.029 | -0.008 | -0.144*** | 0.044 |
| | (0.057) | (0.009) | (0.056) | (0.062) |
| Occ6 | 0.067** | -0.003 | -0.083*** | 0.075** |
| | (0.026) | (0.004) | (0.030) | (0.030) |
| Occ7 | -0.147 | -0.075*** | -1.080*** | -0.033 |
| | (0.412) | (0.016) | (0.127) | (0.454) |
| Constant | 4.484*** | -0.192*** | 1.415*** | 4.340*** |
| | (0.296) | (0.051) | (0.326) | (0.412) |
| $N$ | 5,653 | 5,635 | 5,635 | 5,635 |
| $R^2$ | 0.112 | | | 0.101 |
| IV F-stat | | | | 38.03 |
| Anderson LM | | | | 29.47 |
| CD F-statistic | | | | 14.76 |

Note: The first-stage results are in Columns (2) and (3). Columns (1) and (4) are the OLS and second-stage results. Robust standard errors in parentheses: ***1%, **5%, *10%.

Table 1.15: Quadratic form results of OLS and IV Estimation

| VARIABLES | (1)<br>OLS<br>_R_C<br>lnwage | (2)<br>IV_First<br>_R_C<br>ms | (3)<br>IV_First<br>_R_C<br>ms_sq | (4)<br>IV_First<br>_R_C<br>lnemp_den | (5)<br>IV_Second<br>_R_C<br>lnwage |
|---|---|---|---|---|---|
| lmin_dis | | -0.032*** | -0.096*** | -0.164*** | |
| | | (0.002) | (0.013) | (0.014) | |
| historic | | 0.051*** | -0.004 | -0.034** | |
| | | (0.004) | (0.032) | (0.015) | |
| structure_90 | | -0.020** | 0.145* | 0.723*** | |
| | | (0.010) | (0.077) | (0.091) | |
| lmin_dis_sq | | -0.003*** | 0.005** | -0.030*** | |
| | | (0.000) | (0.002) | (0.002) | |
| structure_90_sq | | -0.012*** | -0.237*** | -0.262*** | |
| | | (0.003) | (0.026) | (0.027) | |
| ms | -0.135 | | | | -0.036 |
| | (0.070) | | | | (0.299) |
| ms_sq | 0.925** | | | | 0.066* |
| | (0.111) | | | | (0.034) |
| lnemp_den | -0.051** | | | | -0.055 |
| | (0.012) | | | | (0.047) |
| educ | 0.065*** | -0.001*** | -0.005 | 0.006** | 0.065*** |
| | (0.006) | (0.000) | (0.003) | (0.002) | (0.003) |
| male | 0.177*** | 0.003 | 0.022 | -0.020 | 0.177*** |
| | (0.014) | (0.002) | (0.016) | (0.013) | (0.013) |
| exp | 0.029*** | -0.001*** | -0.005* | 0.004* | 0.030*** |
| | (0.002) | (0.000) | (0.003) | (0.002) | (0.003) |
| exp_sq | -0.001** | 0.000* | 0.000 | -0.000* | -0.001*** |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| marriage | 0.065 | 0.013*** | 0.098*** | -0.022 | 0.063*** |
| | (0.028) | (0.004) | (0.028) | (0.024) | (0.024) |
| lnpubexp | -0.055 | 0.107*** | 1.280*** | 0.473*** | -0.049 |
| | (0.115) | (0.006) | (0.064) | (0.040) | (0.042) |
| lnexpst | 0.087*** | 0.070*** | 0.453*** | -0.050*** | 0.087*** |
| | (0.006) | (0.002) | (0.011) | (0.012) | (0.017) |
| lnexpedu | 0.107 | 0.025*** | 0.016 | -0.491*** | 0.104*** |
| | (0.111) | (0.005) | (0.035) | (0.044) | (0.037) |
| lnfixedasset | -0.052*** | -0.140*** | -1.188*** | 0.399*** | -0.055* |
| | (0.004) | (0.004) | (0.036) | (0.023) | (0.031) |

Note: ***Significant at 1%; **Significant at 5%; *Significant at 10%.

Table 1.16: Continued Table 1.15

| VARIABLES | (1) OLS _R_C lnwage | (2) IV_First _R_C ms | (3) IV_First _R_C ms_sq | (4) IV_First _R_C lnemp_den | (5) IV_Second _R_C lnwage |
|---|---|---|---|---|---|
| Occ1 | 0.205** | -0.000 | -0.007 | -0.021 | 0.204*** |
|  | (0.025) | (0.006) | (0.049) | (0.042) | (0.042) |
| Occ2 | 0.235*** | -0.004 | -0.036 | -0.074*** | 0.233*** |
|  | (0.006) | (0.004) | (0.030) | (0.027) | (0.026) |
| Occ3 | 0.076* | 0.003 | 0.030 | -0.048* | 0.075*** |
|  | (0.020) | (0.004) | (0.030) | (0.027) | (0.026) |
| Occ4 | -0.051* | -0.009** | -0.060** | -0.085*** | -0.053** |
|  | (0.013) | (0.004) | (0.028) | (0.027) | (0.025) |
| Occ5 | 0.049 | -0.010 | -0.095 | -0.138*** | 0.046 |
|  | (0.080) | (0.009) | (0.066) | (0.050) | (0.063) |
| Occ6 | 0.040** | -0.003 | -0.024 | -0.067** | 0.039 |
|  | (0.009) | (0.004) | (0.028) | (0.027) | (0.025) |
| Occ7 | 0.204* | 0.007 | -0.124 | -0.204 | 0.201 |
|  | (0.053) | (0.028) | (0.177) | (0.170) | (0.207) |
| Constant | 3.310*** | -0.301*** | -5.358*** | -0.321 | 3.341*** |
|  | (0.323) | (0.043) | (0.379) | (0.278) | (0.244) |
| $N$ | 9,086 | 9,061 | 9,061 | 9,061 | 9,061 |
| $R^2$ | 0.268 |  |  |  | 0.267 |
| IV F-stat |  |  |  |  | 149.7 |
| Anderson LM |  |  |  |  | 321.3 |
| CD Wald F-statistic |  |  |  |  | 91.36 |
| Hansen p-value |  |  |  |  | 0.921 |

Note: ***Significant at 1%; **Significant at 5%; *Significant at 10%.

## 1.A.2 Additional Equations

$$argmin \sum_{i=1}^{n} \left( lnwage_{i,c}^{native} - \beta_0 - \delta_1 MS_c - \delta_2 MS_c^2 - \boldsymbol{\beta_1 X_{i,c}} - \boldsymbol{\beta_2 City_c} - \sum_{i=1}^{7} \right)^2 \tag{1.13}$$

we generalize the above formula as the following:

$$Y = \boldsymbol{X\beta} + MS_c^2 \gamma_2 + MS_c \gamma_1 + \epsilon \tag{1.14}$$

where $\boldsymbol{X}$ is a vector including all the independent variables other than the quadratic term in my research. We can thus rewrite the formula in this form:

$$Y = \boldsymbol{Z\delta} + \epsilon \tag{1.15}$$

where $\boldsymbol{Z}$ is a vector that combines all the independent variables and $\boldsymbol{\delta}$ a new vector that combines $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$. Based on the OLS estimation, we can get the estimated results of the coefficients as:

$$\widehat{\delta} = \left( Z^T Z \right)^{-1} Z^T Y \tag{1.16}$$

In this process, we treat the quadratic as a whole and take it as a common linear term in the estimation progress. The estimated results of the coefficients are thus unbiased and consistent. It also makes our estimation process, both the first and second stages, remain the same and the traditional econometric theory is still applicable in this related part. Based on this fact, I add the corresponding quadratic terms of instrument variables as the new instrument variables and rerun the previous codes to get the unbiased coefficients of the quadratic term.

# Chapter 2

# The Impact of Family Background Risks on Financial Risk-asset Allocation

Investigating household finance presents challenges due to the intricate nature and measurement of household behaviours and the presence of constraints not addressed by conventional models. The allocation of financial risk assets by households involves peculiar non-diversifiable risks, termed background risks. Comprehensive knowledge of these background risks aids families in recognising their limitations and avoiding unsuitable investments. This research advances the nuanced understanding of household decision-making in the allocation of financial risk assets, particularly under the influence of household-originated risks. Employing Cragg's Double Hurdle Model and utilising household-level data from the 2017 China Household Finance Survey (CHFS), this study explores the impact of background risks on household allocation of financial risk assets. Empirical findings underscore that background risks exert discernible effects on household participation in the risky market and their holding of risky assets, in line with the literature.

## 2.1  Introduction

Household finance often deviates from the predictions of portfolio theory (Guiso et al., 2002; Campbell, 2006) because of the many specificities that give the field its unique nature. Investors (individuals or households) are often exposed to other sources of risk, such as changes in earned income, real estate or business holdings or values, and unexpected expenses such as those related to health issues, leading to changes in asset allocation (Davis and Willen, 2000; Franke et al., 2006; Baptista, 2008; Edwards, 2008; Cardak and Wilkins, 2009; Curcuru et al., 2010; Palia et al., 2009, 2014). Since Heaton and Lucas (2000), these non-financial market risks are called "background risks", and the assets are affected by the background risks are called "background assets" (Gollier, 2001). Because background assets are illiquid, it is difficult to control background risk by adjusting the holdings of these assets. Investors often focus more on overall risk when investing than just portfolio risk. The empirical evidence suggests that absolute risk aversion decreases, with individuals becoming more willing to take risks as the absolute amount of wealth increases (Gollier and Pratt, 1996; Guiso and Paiella, 2008; Bacon et al., 2020). This implies the phenomenon that individuals become risk-averse when background risks are added to personal wealth (Gollier and Pratt, 1996). Therefore, as background risk may greatly affect individual investment decisions, it is necessary to empirically explore the issue of household financial asset allocation by controlling for this type of risk. Having a better understanding of background risks and efficient allocation of family assets will not only help the growth of family wealth but also help families reasonably reduce their own risks.

In the existing literature, a few researchers worked on portfolio selection, considering background risk. Heaton and Lucas (1997, 2000) use a theoretical model to explore the interaction of background risks, such as labour income and illiquid assets. Cocco et al. (2005) present a realistically calibrated life cycle model of consumption and portfolio choice with labour income and find that it plays a crucial role in investment behaviour as it substitutes for riskless asset holdings. Edwards (2008) develops a theoretical model where health risk determines portfolio shares through its effect on risk aversion. Rosen and Wu (2004); Decker and Schmitz (2015) find a health shock increases individual risk aversion; however, Love and Smith (2010) find health risk may increase household participation. Chetty et al. (2017) find that increases in property value reduce stockholdings, while increases in home equity wealth raise stockholdings. Li et al. (2021) show that entrepreneurship significantly decreases both household risky market participation and risky asset holding.

Overall, the role of background risk in household financial behaviour remains

underexplored. The studies demonstrate that background risk significantly influences household investment portfolio choices and asset allocation behaviour. However, conventional models like mean-variance analysis, binary response models, multiple linear regression, and the Tobit model encounter limitations in explaining non-stockholding and undiversified portfolios. Prior research often presupposes that individuals either engage in the market for risky assets or do not, neglecting the scenario where a household may still participate in financial risk markets without direct participation in stockholding. Cragg's Double Hurdle (CDH) model offers a solution by acknowledging that certain individuals may never participate in such markets due to specific characteristics. This model addresses the shortcomings of previous studies by relaxing the assumption that all households exhibit uniform risk aversion and propensity to invest. The CDH model is particularly adept at managing datasets characterised by a significant presence of zero observations.

This study is the first to investigate the relationship between four types of background risk and household portfolio choices in the Chinese context, utilising a large dataset to provide comprehensive insights from the perspective of background risk. This study aims to answer the following research questions: How do background risks influence the decision to participate in and the extent of investment in financial risk assets among Chinese households? How does the presence of health risks among household members impact the likelihood and extent of financial risk asset allocation? Are there significant differences in financial risk asset allocation between urban and rural households in China? To my knowledge, the CDH model is the first attempt at using it in research in household finance. This article uses data from the 2017 China Household Finance Survey (CHFS) to empirically examine the impact of labour income, health, housing, and private enterprise on household participation in the financial market. Moreover, my paper is the first to comprehensively study these four background risks, which have been separately advocated and studied in the previous literature.

The results are in line with the literature and show that an increase in the number of family members with stable employment positively influences the family's inclination to allocate risky financial assets, with each additional person contributing to a 5.4% increase. Conversely, a higher proportion of family members with poor health leads to a 4.4% reduction. However, I do not find statistical evidence proofing the family holding health insurance correlation. Furthermore, the presence of real estate or business in households corresponds to a decrease in the allocation of risky financial assets by 3.2% and 4.1% respectively. Similarly, an increase in the amount of real estate or commercial assets held by households leads to a respective decrease of 10.7% and 0.3%.

Moreover, I also estimate the optimal stable working force number in Chinese

families, which is around 2, which would give more confidence to allocate assets in the financial risk market. Furthermore, urban samples are more likely to allocate assets in the financial risk market. When the proportion of members with poor health exceeds about 60%, the allocation of financial risk assets will be reduced. Overall, I empirically demonstrate the background risks in determining household asset allocation, helping us understand the impact of background risks and the determinants of household financial asset allocation.

My findings also have implications for policymakers and financial practitioners. Policymakers should implement financial literacy programs, create targeted savings schemes with tiered risk options, and provide subsidised risk-reduction tools like insurance to encourage safer investments. Financial institutions should develop risk-based financial products, use behavioural nudges to guide customers toward safer investments, and offer personalised financial advice to match risk profiles.

The content of this chapter is arranged as follows: Section 2.2 includes a review of the most relevant literature; Section 2.3 describes the database, main variable relations, Section 2.4 the empirical models; Sections 2.5 and 2.6 reports the results and robustness checks; Section 2.8 draws the conclusion.

## 2.2 Literature Review

This summary outlines the research conducted on theoretical models that incorporate background risk and reviews the empirical literature exploring its relevance in household portfolio allocation.

The inclusion of background risk, as initially formalised by Pratt and Zeckhauser (1987), leads households to increase precautionary savings and exhibit risk-averse behaviour in response to risks such as stock market volatility. This behavioural response, known as "temperance", according to Hochguertel (2003), holds true across a wide range of utility specifications, as further discussed in (Heaton and Lucas, 2000). Gollier and Pratt (1996) introduce a novel concept in utility functions called "risk vulnerability" (RV), which assumes that the addition of an unfair background risk to wealth induces more risk-averse behavior towards any other independent risk. They demonstrate that RV is equivalent to the condition that an undesirable risk cannot be made desirable by the presence of an independent, unfair risk. Furthermore, they show that under risk vulnerability, the addition of an unfair background risk reduces the demand for risky assets, justifying deviations from optimal portfolio allocations in the absence of background risk. For instance, Baptista (2008) provides evidence of households reducing their holdings of risky assets due to temperance induced by background risks.

While it is generally expected that an increase in labor income uncertainty would lead to a decrease in the allocation of risky assets, models incorporating labor income risk have yielded diverse results in accurately predicting the avoidance of risky assets (Haliassos and Bertaut, 1995; Heaton and Lucas, 1997, 2000). Heaton and Lucas (2000) employed a calibrated dynamic portfolio allocation model using US data and demonstrated that background risks associated with labor income and proprietary business income can account for some of the observed low stockholding behavior. Cocco et al. (2005) find labour income is similar to implicit holdings of safe assets, stimulating investment in risky financial assets. Fratantoni (1998) through the calibration of US data, found that mortgage commitments and labour income risk induce temperance, resulting in a reduction in risky asset holdings. Campbell and Viceira (2002); Campbell et al. (2003) examined the conditions under which a simplified dynamic model of rational portfolio selection and consumption, incorporating risky labor income and retirement, implies that individuals or households with stable income are more inclined to assume risk and exhibit lower risk aversion. Furthermore, several studies, including (Haliassos and Bertaut, 1995; Jagannathan et al., 1996; Guiso et al., 1996; Bertaut, 1998; Fratantoni, 1998; Elmendorf and Kimball, 2000; Viceira, 2001; Campbell and Viceira, 2002; Storesletten et al., 2004), have revealed a negative impact when families face income risk, crowding out financial investments. The literature highlights labour income risk's impact on portfolio choices. In response, our study uses the CDH model to separately evaluate the likelihood of market participation and the proportion of assets allocated to risky investments. We account for the nuanced effects of labour income risk, particularly how varying heterogeneous households alter their risk profiles and investment behaviours. This method is critical in uncovering hidden patterns in data that standard models might overlook.

Health risk can exert a comparable influence (Campbell, 2006), as it amplifies labour income risk and impacts the investment horizon by influencing life expectancy, despite the possibility of health costs being partially insured. Campbell (2006) suggests that poor health can be perceived as a source of both labour income risk and expenditure risk. Rosen and Wu (2004); Edwards (2008) demonstrates the escalating risk associated with health as age, leading to a decrease in their exposure to financial risk as a hedge. Love and Smith (2010) find health risk may increase household participation. The literature on precautionary saving, along with various articles examining the impact of future health expenditures on saving, has shed light on this relationship (Hubbard et al., 1994; Guiso et al., 1996; Palumbo, 1999; Dynan et al., 2004). Combining the perspective of preventive savings, health risks can be analysed from two perspectives: one is the decline in bad health or temporary health conditions, and the other is the hedging of health risks when the family holds a

sufficient amount of medical insurance. Health risk has been shown to significantly affect financial decisions. In our analysis, we extend this understanding by applying the CDH model to assess how health risks, quantified by poor health ratios and health insurance coverage, affect both financial market participation and the allocation of risky assets. This approach enables us to identify the specific channels through which health concerns impact household behavior, offering a more nuanced understanding of how health risks shape financial strategies.

In addition to income risk and health risk, two other factors warrant attention: home and business risk. Housing represents the most substantial asset for the majority of families, which holds true in China as well. In 2018, the net property value of Chinese urban households accounted for 71.35% of per capita household wealth, while for rural households, it accounted for 52.28% (National Bureau of Statistics of China). Previous literature has predicted that housing tends to reduce the demand for risky assets due to increased household risk and illiquidity (Grossman and Laroque, 1987; Brueckner, 1997; Fratantoni, 2001; Chetty and Szeidl, 2007). However, empirical studies have shown varied systematic relationships between housing and investment portfolios in practice. Houses are durable assets that offer owners various housing services. In this regard, they resemble long-term bonds and can serve as a safeguard against fluctuations in the relative costs of housing and non-housing consumption (Cocco, 2005; Pelizzon and Weber, 2005; Sinai and Souleles, 2005; Dröes and Hassink, 2013; Chetty et al., 2017). However, housing is also an asset that is not easily converted into cash, making it expensive for homeowners to modify their usage of housing services in reaction to economic disturbances. The absence of sufficient liquidity might impede individuals from acquiring and possessing residential properties, as well as assuming financial liabilities. Therefore, considering the heterogeneity of background risk, the presence of a house or the amount of house assets may have inconsistent impacts on future investment decisions and asset allocation.

Moreover, family background risks arising from entrepreneurship or business also influence the decision-making and extent of family participation in the financial risk market. The impact of business on household financial asset allocation can be understood through two channels. First, as described by Chetty et al. (2017), an increase in entrepreneurial assets leads to a decrease in effective risk aversion towards liquid wealth, prompting business families to invest more in risky financial assets to diversify their venture capital portfolios. Second, the risk substitution effect, related to business background risk, implies that households involved in business may reduce their financial asset holdings to limit total risk exposure. This is due to entrepreneurship being a dynamic and uncertain endeavour with high background risk (Hochguertel, 2003). Among others, Heaton and Lucas (2000); Palia et al.

(2014); Ghahreman (2016) demonstrate that entrepreneurial risk would decrease financial asset holding. However, the specific financial asset allocation behavior of business households in China's household-dependent financing system and whether entrepreneurship affects household financial behavior through other channels beyond entrepreneurial risk remains unclear. Although Li et al. (2021) has demonstrated that entrepreneurship significantly reduces household risk market participation and risk asset holdings in China, more empirical data are needed to examine this relationship while considering different channels. The dual nature of housing and business risks poses unique challenges to financial decision-making. Our study addresses these challenges by leveraging the CDH model, which allows us to disentangle the effects of housing and business ownership on market participation and asset allocation. By incorporating this model, we capture the differential impact of housing and business risks on financial behavior, providing insights into how these background risks lead to significant deviations from optimal portfolio allocation, particularly in the Chinese context.

By utilizing the Chinese Household Finance Survey (CHFS) 2017, I employ Cragg's Double Hurdle Model (CDH Model) to investigate the influencing factors of background risk on the risk investment allocation of household financial assets. This comprehensive approach will enable a deeper understanding of the current risk of investment and asset allocation behaviour among Chinese households.

## 2.3 Data and Summary Statistics

### 2.3.1 Data Description

In this paper, I take variables from 2017 China Household Finance Survey (CHFS) which is a cross-sectional nationwide survey conducted by the China Household Finance Survey and Research Center. The main content includes demographic characteristics and employment, assets and liabilities, income and consumption, social security and insurance, subjective attitudes, and other related information, which comprehensively and meticulously portrays family economic and financial behaviours. The nature of this dataset makes it possible to distinguish financial assets into risky financial assets and risk-free financial assets. The risky financial assets include stocks, stock funds, bond funds, hybrid funds, financial bonds, corporate bonds, financial wealth management products, derivatives, non-RMB assets, and other financial assets. Risk-free financial assets include time deposits, demand deposits, local government bonds, monetary funds, treasury bills, cash, and loans.

As in Cai et al. (2013), the dependent variables are defined as follows: 1) Households that hold one of the financial risk asset(s) (or more), are considered as participation samples and denoted by $risk\_if$ as 1, otherwise; 2) the household's risky financial asset allocation is calculated as the proportion of risky financial assets within the overall household financial portfolio. This proportion is denoted as $riskratio$ and falls within the range of $[0, 1]$.

To comprehensively investigate the influence of background risk on the allocation of financial risk assets in households, this study would include the following three perspectives labour income risk, health risk, housing risk, and business risk.

Regarding labour income risk, simply summarising the labour income of the respondents may not represent the labour income risk of the entire family, as in Heaton and Lucas (2000); Cocco (2005); Campbell (2006); therefore, I use the number of individuals employed within the household as a proxy[1] This variable is a proxy of labour income risk, as households with more employed individuals may face varying levels of income uncertainty due to fluctuations in employment status and wages. $labnum$ is defined as the number of the respondent's family members employed. The largest number of stable workers in a household is 6. If no family member of the respondent's family is employed by others or units and has signed a formal contract, it takes value 0, and it is set as the control group.

Respect to health risks, we consider two dimensions following Rosen and Wu (2004); Edwards (2008): health status and medical insurance, which are defined as the proportion of individuals with poor health in the household and the extent of medical insurance coverage representing the health risk[2], respectively. Worsening health risks could lead households to prepare more precautionary savings and reduce financial asset allocation. First, health status is proxied by when family members answered "average", and "bad" to their current physical conditions, they were considered to have poor health status, then divide the number of poor health condition family members by the number of family members to get the ratio of the poor health status of family members, $health\_bad\_ratio$. Secondly, the family medical insurance coverage rate $health\_insur\_ratio$ is calculated, that is, the number of family

---

[1]While the number of workers in a household serves as an inverse proxy for labor risk, this measure operates under the assumption that household members are employed in diverse occupations, thereby mitigating the risk of job loss through occupational diversification. However, it is plausible that all household members are employed within the same occupation or industry (e.g., in the same factory), which would effectively concentrate and amplify labor risk rather than diversifying it. Unfortunately, due to data limitations, this critical aspect cannot be directly addressed or accounted for in the present analysis.

[2]Medical insurance includes basic medical insurance for urban and rural residents, basic medical insurance for urban employees, rural cooperative medical insurance, public medical care, commercial medical insurance purchased by units, commercial medical insurance purchased by individuals, and supplementary medical insurance by enterprises.

members participating in medical insurance accounts for the total number of family members.

The heterogeneity of housing risk leads to different conclusions in the literature, it either has a "crowding out effect", see Fratantoni (2001); Chetty and Szeidl (2007), or as in Cocco (2005); Goetzmann and Valaitis (2006); Chetty et al. (2017) have reached the opposite conclusion. The main reason is that the foreign mortgage mechanism is mature, investors can mortgage their own real estate to obtain funds to invest in the risky financial market. This paper measures the impact of housing risk on China's household financial risky asset allocation in two dimensions: whether households own residential properties and the proportion of housing assets within the family's overall asset portfolio. The dummy variable for whether to maintain own property is denoted by *house_if*. If the answer is "yes", it is 1, otherwise. Set *house_per* as the ratio of the current total value of owned property to the total household assets. The intuition is that households that own residential properties may face risks associated with changes in property values, housing market volatility, and the potential illiquidity of housing assets.

Lastly, in terms of business risk, Heaton and Lucas (2000) demonstrate that households owning private companies tend to hold a lower proportion of stocks within their financial wealth, despite having similar financial wealth to others. Therefore, this paper sets up two variables to measure whether households own businesses and the value of the family's business assets. 1) Set the dummy variable of whether to maintain business assets as *business_if*, if the answer is "yes", the value is 1, otherwise; 2) set *business_asset* as the current total value of business assets in the household. The intuition is that households involved in business ownership may face risks related to business performance, economic conditions, and the uncertain outcomes of entrepreneurial endeavours.

This study makes use of several control variables to account for family characteristics that could influence financial risk market participation and asset allocation decisions. First, the study considers the proportion of household allocation of financial assets, which is influenced by both wealth and income effects. The variables *ln_ave_income* and *ln_tot_asset* are used to represent the annual disposable income and total assets of a household, respectively. These factors are likely to affect a household's decision to engage in the financial risk market (Jie and Mengyuan, 2018).

Additionally, risk attitude plays a significant role in household financial decisions, particularly in asset allocation (Zhang et al., 2021). Suppose the total risk faced by investors is fixed, the proportion of investors' participation in risky financial markets and the allocation of risky financial assets are determined by their risk attitudes

(Charness and Gneezy, 2010; Siebenmorgen and Weber, 2003). Risk-averse investors are less willing to allocate more risky financial assets, risk-averse financial assets are relatively low, and even risk-averse investors do not participate in risky financial markets at all. Household risk attitudes are captured through a series of dummy variables. If households are considered to have a preference for risky assets respondents indicate a preference for "high-risk, high-reward items" or "slightly higher-risk, slightly higher-reward items" in the questionnaire, which is captured by the variable *attitu_good*. Households expressing a preference for "average risk, average return" are considered to have a neutral risk preference, denoted by *attitu_ok*. Conversely, those who prefer "slightly lower risk, slightly lower return items" or are "unwilling to take any risk" are classified as risk-averse, represented by *attitu_low*. For respondents who select "I don't know," their risk attitude remains undefined.

Another crucial factor is financial knowledge, which influences a household's ability to navigate the risky financial market. As suggested by Wang (2009), this study evaluates the usual attention respondents give to economic and financial information. Households that are "very concerned" about such information are considered to have a strong willingness to learn about financial knowledge, captured by *knowledge_good*. Those who are "general" in their attention are denoted by *knowledge_ok*, indicating occasional financial knowledge. In contrast, households that "rarely pay attention" or "never pay attention" are considered to lack financial knowledge, indicated by *knowledge_low*.

Furthermore, the study accounts for urban and rural household registration, recognising that urban households generally have more exposure to risky financial markets due to differences in cognition and living environments. The dummy variable *rural* is employed, where a value of 1 represents rural and 0 represents urban.

Lastly, education level is another critical control variable, as families with higher educational attainment are more likely to understand and rationally analyse the risks associated with financial products. This study uses the highest educational attainment within the household to measure this variable. When the highest level of education is no-schooling, primary school, or junior high school, the variable *edu_mid* is set to 1. If the highest level of education is high school, technical secondary school, or vocational high school, the variable *edu_high* is set to 1. For households where the highest education is a college or bachelor's degree, *edu_college* takes the value of 1, while for those with a master's or doctoral degree, *edu_graduate* is set to 1.

Finally, this study also controls for family size, marriage, and household age which is represented by the age of the head of the household and its quadratic term in the control variable group.

Table 2.1 reports the independent variables' variable meaning and their symbolic features. Table 2.2 reports the control variables' variable meaning and their symbolic features.

Table 2.1: Independent Variables Definition

| Independent Variable Setting Description | | |
|---|---|---|
| Independent Variables | Variable Meaning | Symbolic Features |
| Labor income risk | Number of family members with fixed income | **labnum** |
| Health risk | Self-evaluation of the health status of family members | **health_bad_ratio** |
| | Home health insurance coverage | **health_insur_ratio** |
| Housing risk | Housing (*whether having ownership*) | **house_if** |
| | Proportion of the house value taking in the family's total asset | **house_per** |
| Business risk | Business assets (*whether having the ownership*) | **business_if** |
| | Value of the business | **business_asset** |

Table 2.2: Control Variables Definition

| Control Variable Setting Description | | |
|---|---|---|
| Control Variables | Variable Meaning | Symbolic Features |
| Total asset and income | Log of total asset value in a household | **ln_tot_asset** |
| | Log of average annual disposable income of family members | **ln_ave_income** |
| Risk attitude | Family's appetite for risk | **attitu_good, attitu_ok, attitu_low** |
| Financial knowledge | Attention to economic and financial information | **knowledge_good, knowledge_ok, knowledge_low** |
| Urban-Rural Hukou | Hukou location | **rural** |
| Education level | The most educated member of the family | **edu_high, edu_colleage, edu_graduate** |
| Population size | Family size | **number** |
| Age | Hoseuhold's head age and its quadratic term | **age, age_sq** |
| Marriage | Whether got married | **married** |

## 2.3.2 Summary Statistics

There are 35,626 remaining observations. 5,137 households participated in risky financial assets, accounting for 14.4%. Financial assets accounted for an average of 6.34% of total assets. Tables 2.3 and 2.4 reports descriptive statistics for the full sample and urban-rural, respectively. From Table 2.3 we can see that in a family, the healthy insurance cover rate is 90.2%, 84.3% families own real estate, and the house value is worth about 59.1% in overall family asset. However, only 13.0% families own a self-business, and it is worth 10.9% family assets. From Table 2.4, we can see that the urban sample is more likely to participate and address more family assets in financial risky market. By comparing their standard deviations, there is more variability in the dependent variables for the urban sample.

Table 2.3: Summary Statistics of the Main Variables (Full Sample)

| Variable | Obs | Mean | SD | Min | Max |
|---|---|---|---|---|---|
| **Dependent Variables** | | | | | |
| risk_if | 35,623 | 0.144 | 0.351 | 0 | 1 |
| riskratio | 35,623 | 0.040 | 0.127 | 0 | 1 |
| riskratio> 0 | 5,137 | 0.274 | 0.217 | 0 | 1 |
| **Explanatory Variables** | | | | | |
| labnum | 35,623 | 1.449 | 0.612 | 0 | 6 |
| health_bad_ratio | 35,623 | 0.457 | 0.401 | 0 | 1 |
| health_insur_ratio | 35,623 | 0.902 | 0.233 | 0 | 1 |
| house_if | 35,623 | 0.843 | 0.364 | 0 | 1 |
| house_per | 35,623 | 0.591 | 0.406 | 0 | 1 |
| business_if | 35,623 | 0.130 | 0.337 | 0 | 1 |
| business_asset | 35,623 | 0.109 | 1.716 | 0 | 200 |
| **Control Variables** | | | | | |
| tot_asset | 35,623 | 0.954 | 2.496 | 0 | 208.251 |
| edu_high | 35,623 | 0.166 | 0.372 | 0 | 1 |
| edu_college | 35,623 | 0.342 | 0.406 | 0 | 1 |
| edu_graduate | 35,623 | 0.032 | 0.124 | 0 | 1 |
| attitu_good | 35,623 | 0.036 | 0.186 | 0 | 1 |
| attitu_ok | 35,623 | 0.068 | 0.253 | 0 | 1 |
| attitu_low | 35,623 | 0.166 | 0.372 | 0 | 1 |
| knowledge_good | 35,623 | 0.032 | 0.175 | 0 | 1 |
| knowledge_ok | 35,623 | 0.058 | 0.234 | 0 | 1 |
| married | 35,623 | 0.830 | 0.375 | 0 | 1 |
| age | 35,623 | 44.714 | 10.665 | 16 | 60 |
| rural | 35,623 | 0.513 | 0.500 | 0 | 1 |
| number | 35,623 | 3.160 | 1.529 | 1 | 15 |
| ave_income | 35,623 | 0.012 | 0.022 | 0 | 0.818 |

Note: The scale of family average income, business asset and total asset, due to visibility and readability, are divided by 100,000.

Figure 2.1: Distribution positive value of variable *riskratio*
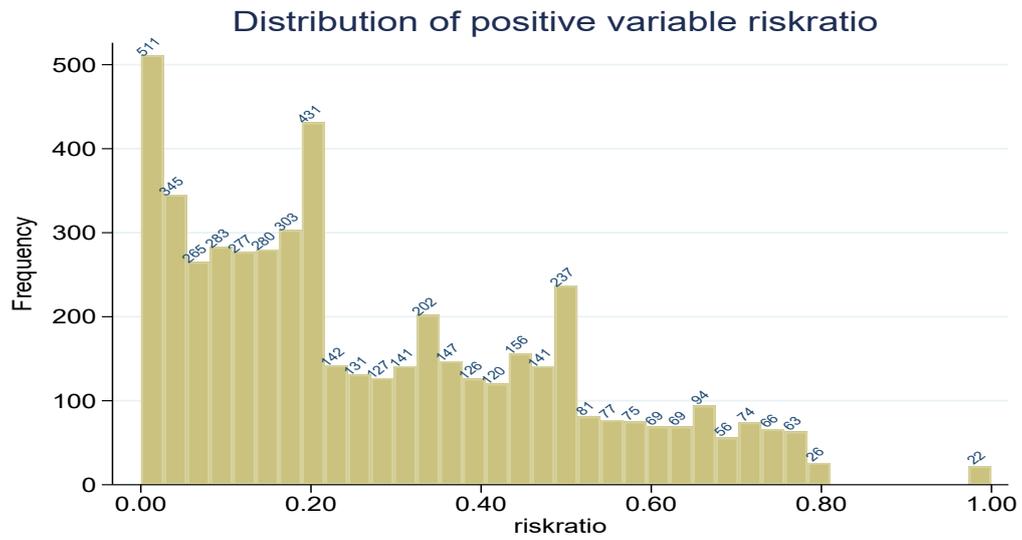
Table 2.4: Summary statistics of the main variables (Urban and Rural Samples)

| Variable | Urban Sample | | | | | Rural Sample | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Obs | Mean | SD | Min | Max | Obs | Mean | SD | Min | Max |
| **Dependent Variables** | | | | | | | | | | |
| risk_if | 17,344 | 0.236 | 0.425 | 0 | 1 | 18,279 | 0.057 | 0.231 | 0 | 1 |
| riskratio | 17,344 | 0.067 | 0.161 | 0 | 1 | 18,279 | 0.013 | 0.072 | 0 | 1 |
| riskratio> 0 | 4,101 | 0.285 | 0.219 | 0 | 1 | 1,036 | 0.230 | 0.202 | 0 | 1 |
| **Explanatory Variables** | | | | | | | | | | |
| labnum | 17,344 | 0.692 | 0.851 | 0 | 5 | 18,279 | 0.359 | 0.681 | 0 | 6 |
| health_bad_ratio | 17,344 | 0.425 | 0.404 | 0 | 1 | 18,279 | 0.489 | 0.395 | 0 | 1 |
| health_insur_ratio | 17,344 | 0.900 | 0.226 | 0 | 1 | 18,279 | 0.904 | 0.239 | 0 | 1 |
| house_if | 17,344 | 0.835 | 0.371 | 0 | 1 | 18,279 | 0.850 | 0.357 | 0 | 1 |
| house_per | 17,344 | 0.645 | 0.393 | 0 | 1 | 18,279 | 0.539 | 0.411 | 0 | 1 |
| business_if | 17,344 | 0.119 | 0.324 | 0 | 1 | 18,279 | 0.141 | 0.348 | 0 | 1 |
| business_asset | 17,344 | 0.136 | 2.222 | 0 | 200 | 18,279 | 0.083 | 1.028 | 0 | 65.25 |
| **Control Variables** | | | | | | | | | | |
| tot_asset | 17,344 | 1.498 | 3.173 | 0 | 208.251 | 18,279 | 0.437 | 1.430 | 0 | 68.520 |
| edu_high | 17,344 | 0.197 | 0.398 | 0 | 1 | 18,279 | 0.136 | 0.343 | 0 | 1 |
| edu_college | 17,344 | 0.306 | 0.461 | 0 | 1 | 18,279 | 0.114 | 0.318 | 0 | 1 |
| edu_graduate | 17,344 | 0.026 | 0.159 | 0 | 1 | 18,279 | 0.006 | 0.074 | 0 | 1 |
| attitu_good | 17,344 | 0.048 | 0.214 | 0 | 1 | 18,279 | 0.025 | 0.155 | 0 | 1 |
| attitu_ok | 17,344 | 0.084 | 0.278 | 0 | 1 | 18,279 | 0.053 | 0.225 | 0 | 1 |
| attitu_low | 17,344 | 0.207 | 0.405 | 0 | 1 | 18,279 | 0.128 | 0.334 | 0 | 1 |
| knowledge_good | 17,344 | 0.048 | 0.213 | 0 | 1 | 18,279 | 0.017 | 0.128 | 0 | 1 |
| knowledge_ok | 17,344 | 0.083 | 0.275 | 0 | 1 | 18,279 | 0.035 | 0.184 | 0 | 1 |
| ave_income | 17,344 | 0.016 | 0.027 | 0 | 0.818 | 18,279 | 0.009 | 0.013 | 0 | 0.29 |
| married | 17,344 | 0.806 | 0.395 | 0 | 1 | 18,279 | 0.853 | 0.354 | 0 | 1 |
| age | 17,344 | 44.350 | 10.854 | 16 | 60 | 18,279 | 45.035 | 10.486 | 16 | 60 |
| number | 17,344 | 2.829 | 1.272 | 1 | 15 | 18,279 | 3.473 | 1.679 | 1 | 15 |

Note: The scale of family average income, business asset and total asset, due to visibility and readability, are divided by 100,000.

## 2.4  Methodology

In this section, I discuss the methodology implemented in this study. The core purpose of this study is to examine the impact of household background risk on household risky financial asset allocation. Certain households may opt out of participating in financially precarious endeavours, resulting in the dependent variables having a value of zero. To analyze this, the standard Tobit model, as outlined by Greene (2003), is suitable. This model allows for the investigation of factors influencing both the decision to participate and the degree of involvement. It has been employed in household finance empirical studies, such as Lang and Wu (2014); West and Worthington (2017); Huang and Lu (2021); Khan et al. (2022). This model presents some limitations when the proportion of zero values in the dependent variable is large because the parameterisation of the Tobit model is strict. Tobit model only considers the dependent variable censored to zero and ignores the sources and heterogeneity of zero observations (Newman et al., 2003; Martínez-Espiñeira, 2006).

To thoroughly investigate the influence of family background risks on households' financial risk asset allocation, it is crucial to account for two distinct types of families who do not participate in the risky financial market. Type-I families comprise those who are willing to participate in the market but face constraints that prevent them from doing it, such as insufficient principal or limited access to financial resources. On the other hand, Type-II families encompass those who have no intention of engaging in the financial market, regardless of the amount of principal they possess. By considering these two types of families separately, we can gain valuable insights into the complex dynamics between family background risks and financial decision-making in households. This motivates the use of the CDH model.

Cragg (1971) introduces the double-hurdle model as an extension of the Tobit model, particularly for analysing household durable expenditures. Unlike the Tobit model, the double-hurdle model has two hurdles: the first hurdle allows for the possibility that a factor or factors may affect both the likelihood of acquisition and the quantity acquired; the second hurdle allows for the analysis of factors that influence the quantity or intensity of the outcome, conditional on the decision to participate. This model accommodates potential zero values in the second stage, which is significant for capturing instances where individuals or households do not participate. In the context of the double-hurdle model, the participation decision is typically modeled using a Probit estimator. The double-hurdle model enables the separation of the decision to participate in the risky financial market from the level of allocation, treating these aspects independently. The double-hurdle model has

found applications in various research, such as Jones (1989); Burton et al. (1994); Dionne et al. (1996); Moffatt (2005); Mutlu and Gracia (2006); Wang (2016*a*).

To the best of our knowledge, this study represents the first application of the CDH model to investigate the influence of family background risks on the allocation of risky financial assets. The two processes of the CDH model are as follows: the first hurdle is the participation decision, which is defined as whether the households participated in the financial risk asset market; the second hurdle, given that the participation decision is greater than zero, is the proportion of financial risky assets in a household that the family would like to allocate. There is no fixed order between the two processes, i.e., two error terms are independent of each other (Cragg, 1971; Yen and Su, 1995; Mutlu and Gracia, 2006; Engel and Moffatt, 2014; Eakins, 2016).

The underlying assumption in both models sets the lower limit at zero, reflecting the notion that households in my sample are unlikely to borrow money and invest in the risky asset market. This assumption aligns with China's financial reality for two main reasons. First, loans from banks are prohibited from flooding into security markets. Chinese banks are strictly controlled by the China Banking Regulatory Commission, which stipulates clearly that all loans are forbidden to be used to invest in security markets and that every cash flow is strictly censored by the banks. If the borrower tries to invest in the security market with loans, the bank will immediately require a repayment. This makes it very difficult to invest in the security market with bank loans. Secondly, investment returns in the Chinese securities market have been relatively low over the long term. For instance, the Shanghai Stock Index, one of the most comprehensive indicators, reached 3,000 points in 2007, and by the end of 2023, it remained around the same level.

### 2.4.1 The Standard Tobit

$$riskratio_i^* = \boldsymbol{\beta}\boldsymbol{X_i} + \varepsilon_i$$
$$i = 1, 2, ..., n \tag{2.1}$$
$$\varepsilon_i \sim N(0, \sigma^2)$$

$$riskratio_i = \begin{cases} riskratio_i^* & \text{if } riskratio_i^* > 0 \\ 0 & else \end{cases}$$

where $riskratio_i^*$ is a latent variable representing individual household $i$'s investment level of the ratio of family asset allocation in the financial risk market, and $riskratio_i^*$ is corresponding the actual ratio of investment in the financial risky market. $\boldsymbol{X_i}$ is a vector of background risks and control variables (see in Tables 2.1 and 2.2). $\boldsymbol{\beta}$ is a

corresponding vector of parameters to be estimated. The standard Tobit model is estimated using maximum likelihood methods. The log-likelihood function for this estimation is:

$$\ell(Tobit) = \sum_0 ln \left[ 1 - \frac{\Phi(x_i\beta)}{\sigma} \right] + \sum_+ ln \left[ \frac{1}{\sigma}\phi\frac{riskratio_i - x_i^*\beta}{\sigma} \right] \tag{2.2}$$

where "$\sum_0$" indicates summation over the zero observations in the sample and "$\sum_+$" indicates summation over positive observations (i.e., $riskratio_i = 0$ and $riskratio_i > 0$). Further, $\Phi(\cdot)$ and $\phi(\cdot)$ are the standard normal distribution and density functions (CDF and PDF), respectively.

## 2.4.2 Cragg's Double-Hurdle Model

The first hurdle, which is the participation equation, as in Equation (2.3) below,

$$risk\_if_i^* = \boldsymbol{\alpha}\boldsymbol{X_i} + \gamma_{1i} \tag{2.3}$$

In the first stage, the binary dependent variable $risk\_if_i$ is dependent on the latent variable $risk\_if_i^*$.

$$risk\_if_i = \begin{cases} risk\_if_i^* & \text{if } risk\_if_i^* > 0 \\ 0 & else \end{cases} \tag{2.4}$$

where $risk\_if_i^*$ is a latent variable describing the participation decision, the value of which is 1 if the household participated and 0 otherwise. $\boldsymbol{X_i}$ is a vector of background risks and control variables (see in Tables 2.1 and 2.2). $\boldsymbol{\alpha}$ is a vector of corresponding coefficients. $\gamma_{1i}$ is an error term that has a normal distribution.

The second hurdle, which is observed in the ratio of financial asset investment, consists of Equation (2.5),

$$riskratio_i^{**} = \boldsymbol{\beta}\boldsymbol{X_i} + \gamma_{2i} \tag{2.5}$$

$$riskratio_i = \begin{cases} riskratio_i^* & \text{if } riskratio_i^{**} > 0, \ risk\_if_i^* > 0 \\ 0 & otherwise \end{cases}$$

where $riskratio_i$ is the revealed proportion of financial risk assets in household's total asset for individual household$_i$, $riskratio_i^*$ is the corresponding latent value of individual i's actual financial risk asset ratio, $\boldsymbol{X_i}$ is a vector of the background risks and control

variables, see in Tables 2.1 and 2.2. $\boldsymbol{\beta}$ is a vector of corresponding coefficients, $\gamma_{2i}$ is an error term that has a normal distribution with the mean of 0 and variance of $\sigma^2$. The variance-covariance matrix is a diagonal matrix that implies that the two error terms are assumed to be independently distributed[3] (Engel and Moffatt, 2014), it can be written as in Equation (2.6). Unobserved variations in financial literacy or exposure to financial crises may influence both the willingness and ability to invest in risky assets. Therefore, my model controls for risk attitude, financial knowledge, and educational level to minimize dependency issues while maintaining the assumption of independent errors.

$$\begin{pmatrix} \gamma_{1i} \\ \gamma_{2i} \end{pmatrix} \sim \boldsymbol{N} \left[ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & \sigma^2 \end{pmatrix} \right] \tag{2.6}$$

Furthermore, in the second stage, the dependent variable is zero or positive, to which the Tobit model applies,

$$riskratio_i^* = \max\{0, riskratio_i^{**}\}$$

Since observed samples are always larger or equal to zero. Finally, the observed variable $riskratio_i$,

$$riskratio_i = risk\_if_i^* \times riskratio_i^* \tag{2.7}$$

Using maximum likelihood techniques, the log-likelihood for the Cragg's Double Hurdle (CDH) model is calculated as follows:

$$\ell(CDH) = \sum_0 ln \left[ 1 - \Phi(x_i\alpha)\Phi\left(\frac{x_i\beta}{\sigma}\right) \right] + \sum_+ ln \left[ \Phi(x_i\alpha)\frac{1}{\sigma}\phi\left(\frac{y_i - x_i\beta}{\sigma}\right) \right] \tag{2.8}$$

where $\Phi(.)$ indicates normal standard cumulative density function ($CDF$), $\phi(.)$ indicates probability density function ($PDF$). $\Phi(x_i\alpha)$ reflects the probability of households participating in the financial-risky assets market by the first hurdle (*Probit* model). Under the assumption of independence, homoscedasticity, and normal distribution between two error terms, the log-likelihood function of the CDH model can be expressed as the summation of the log-likelihood functions of the *Probit* model and the truncated regression model (McDowell, 2003; Aristei et al., 2008). To maximize the log-likelihood function, the two components can be separately maximized (Jones, 1989; McDowell, 2003).

---

[3]This assumption plays a pivotal role in ensuring the consistent estimation of model parameters. Smith (2003) provides an in-depth discussion on the dependency in double-hurdle models, noting that while the assumption of independence between the two stages is commonly adopted for analytical convenience, empirical evidence often does not substantiate the existence of such independence. This discrepancy underscores the need for caution in interpreting the results and highlights the potential for misestimating relationships if dependency between the hurdles is ignored.

## 2.4.3   CDH Model Specification Issues: Heteroskedasticity

The consistency of standard error estimates for Cragg's Double-Hurdle Model relies on the assumptions of homoscedasticity of $\boldsymbol{\gamma_{1i}}$ and $\boldsymbol{\gamma_{2i}}$. However, when dealing with data that have a cross-sectional structure, the homoscedastic assumption is likely to be violated. When heteroscedasticity is present in a limited dependent variable model, it can lead to inconsistent standard error parameter estimates (Maddala and Nelson, 1975; Arabmazar and Schmidt, 1981; Lin and Schmidt, 1984).

I first do robust standard errors to control for the unconditional heteroskedasticity. To address the conditional heteroskedasticity issue, the standard deviation of the error term is allowed to vary across observations by specifying it as a function of a set of exogenous variables (Newman et al., 2003; Aristei and Pieroni, 2008). The model can be modified to allow for heteroscedasticity by specifying the variance of the errors as a function of a set of continuous variables, as follows:

$$\sigma_i = exp\left(\boldsymbol{w_i}\boldsymbol{\theta}\right) \tag{2.9}$$

where $\boldsymbol{w_i}$ is the vector of the variables that affect the conditional variance and represents the variables in $\boldsymbol{x_i}$, $\boldsymbol{w_i} \in \boldsymbol{x_i}$. As to which variables contribute to heteroskedasticity. The diagonal of the covariance matrix will be changed as

$$\begin{pmatrix}\gamma_{1i}\\\gamma_{2i}\end{pmatrix} \sim \boldsymbol{N}\left[\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1 & 0\\0 & \sigma_i^2\end{pmatrix}\right] \tag{2.10}$$

The exponential specification is chosen as it imposes that the standard deviation $\sigma_i$ be strictly positive (Su and Yen, 1996; Yen, 1993; Yen and Jensen, 1996; Newman et al., 2003). The estimation of the related likelihood functions is done by maximum likelihood, and Equation (2.8) will be modified as:

$$\ell(CDH) = \sum_0 ln\left[1 - \Phi(x_i\alpha)\Phi\left(\frac{x_i\beta}{\sigma_i}\right)\right] + \sum_+ ln\left[\Phi(x_i\alpha)\frac{1}{\sigma_i}\phi\left(\frac{y_i - x_i\beta}{\sigma_i}\right)\right] \tag{2.11}$$

I investigate whether there exists a conditional variance problem in the independent variables. Logically, controlling for conditional variance can potentially help mitigate the issue of heteroscedasticity. By including variables in the specification that capture and specify the variables in the error-variance function of the CDH model. I would be able to account for the systematic relationship between the variability of the dependent variable and the independent variables. This can improve the model's ability to capture and explain the variations in the outcome variable, potentially reducing heteroscedasticity. Controlling for conditional variance can indirectly address heteroscedasticity by reducing the unexplained variability in the model. Therefore, this can be done by specifying the variables that affect the

conditional variance in the regression. The basic construction of the detection of conditional variance is to group all independent variables according to the 50 percentile[4], and then perform statistical detection on the variance of the two groups.

### 2.4.4 Marginal Effects

Furthermore, the magnitude of the ML estimates in the Double-Hurdle Model and the standard Tobit cannot be interpreted in the same way as the OLS estimates. However, the sign of the estimates can be interpreted and can present a clearer intuitive interpretation of the factors determining household financial risky asset investment over single equation estimation models such as the Tobit.

The focus of economic interpretation often centres on analysing the marginal effects of the explanatory variables on the expected value of the dependent variable $riskratio_i$ (Jones and Yen, 2000). This analysis provides valuable insights into the impact of the explanatory variables on the expected value of the dependent variable within the context of the double-hurdle model.

## 2.5 Results

In this section, I present and discuss the empirical results. Before delving into the examination of the outcomes presented in Table 2.5, I will explain the structure of Table 2.5. The estimations presented in Table 2.5, specifically in columns (3), (6), and (8), correspond to the parameter $\boldsymbol{\alpha}$, whereas columns (2), (5), and (7) are associated with $\boldsymbol{\beta}$, as delineated in Equations (2.3) and (2.5). These coefficients within the vector illustrate their effects on households' decisions regarding participation and expenditure, respectively. The baseline findings, as showcased in columns (2) and (3), represent the coefficients of the second hurdle and first hurdle within the CDH model, while column (1) pertains to the standard Tobit coefficient.

As mentioned in the Section 2.4.3, in cross-sectional research, the homoscedastic assumption is unlikely held. It is pointed out by the White test, which the null hypothesis is homoskedasticity. The result is shown with a chi-square at 33.9244 and p-value at 0.000, which means it rejects the null hypothesis and makes the estimation less efficient. Therefore, I first use the robust standard error to control it, in Table 2.5 columns (4), (5), and (6) show the results of the CDH model and Standard Tobit after

---

[4]I address the potential issue of heteroscedasticity by controlling for conditional variance in the independent variables. To detect conditional variance, each independent variable is divided into two groups based on its median (50th percentile). I then perform a statistical test to compare the variances between these two groups. This approach helps to capture the systematic relationship between the variability of the dependent variable and the independent variables, thereby improving the model's ability to explain variations in the outcome variable and potentially reducing heteroscedasticity.

Table 2.5: The Results of Cragg's Double-Hurdle and Standard Tobit Model at Coefficient Level

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
|---|---|---|---|---|---|---|---|---|---|
| | Tobit_Ori | CDH_Ori | | Tobit_Robust | CDH_Robust | | CDH_Het | | |
| VARIABLES | riskratio | riskratio | selection_ll | riskratio | riskratio | selection_ll | riskratio | selection_ll | lnsigma |
| labnum | 0.071*** | 0.106*** | 0.127*** | 0.071*** | 0.106*** | 0.127*** | 0.114*** | 0.120*** | 0.168*** |
| | (0.009) | (0.032) | (0.016) | (0.009) | (0.034) | (0.016) | (0.030) | (0.015) | (0.008) |
| health_bad_ratio | -0.010 | 0.031 | -0.008 | -0.010 | 0.031 | -0.008 | 0.020 | 0.112*** | 0.212*** |
| | (0.016) | (0.055) | (0.028) | (0.016) | (0.053) | (0.028) | (0.050) | (0.035) | (0.022) |
| health_insur_ratio | 0.052* | -0.025 | 0.111** | 0.052* | -0.025 | 0.111** | -0.054 | 0.110** | |
| | (0.027) | (0.095) | (0.046) | (0.028) | (0.102) | (0.048) | (0.081) | (0.044) | |
| house_if | -0.116*** | -0.071 | -0.206*** | -0.116*** | -0.071 | -0.206*** | -0.127** | -0.197*** | |
| | (0.018) | (0.059) | (0.030) | (0.018) | (0.059) | (0.031) | (0.052) | (0.030) | |
| house_per | -0.666*** | 0.068 | -1.166*** | -0.666*** | 0.068 | -1.166*** | 0.121* | -1.114*** | |
| | (0.023) | (0.078) | (0.037) | (0.022) | (0.084) | (0.039) | (0.069) | (0.039) | |
| business_if | -0.203*** | -0.062 | -0.332*** | -0.203*** | -0.062 | -0.332*** | -0.082 | -0.327*** | |
| | (0.018) | (0.061) | (0.031) | (0.020) | (0.063) | (0.033) | (0.052) | (0.030) | |
| business_asset | -0.016*** | 0.001 | -0.027*** | -0.016*** | 0.001 | -0.027*** | -0.001 | -0.025*** | |
| | (0.002) | (0.006) | (0.004) | (0.007) | (0.006) | (0.010) | (0.004) | (0.003) | |
| ln_tot_asset | 0.229*** | 0.016 | 0.399*** | 0.229*** | 0.016 | 0.399*** | 0.032** | 0.405*** | -0.056*** |
| | (0.005) | (0.017) | (0.008) | (0.005) | (0.019) | (0.009) | (0.016) | (0.012) | (0.005) |
| Control Variable | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| $N$ | 35,623 | 35,623 | 35,623 | 35,623 | 35,623 | 35,623 | 35,623 | 35,623 | 35,623 |
| pseudo $R^2$ | 0.287 | 0.394 | | 0.287 | 0.394 | | 0.699 | | |

Note: All estimations include control variables because the reported results table would exceed the page length if it includes control variables; therefore, the full table is reported in Table 2.15 and Table 2.16. ***Significant at 1%; **Significant at 5%; *Significant at 10%.

robust standard errors. The results show that before and after a robust standard error procedure, it only relaxes the standard errors but does not change the value of the estimators and significance.

In Table 2.5, column (7)-(8) are the CDH result after controlling conditional variance, which are from variables *labnum*, *health_bad_ratio*, and *ln_tot_asset*.[5]. The coefficients on *labnum*, *health_bad_ratio*, and *ln_tot_asset* are strongly significant at 1% level in column (9), which means they affect the conditional variance of the outcome model. We can find evidence that in this specification, the standard error would be smaller after the robust standard error procedure, by comparing columns (4) and (7).

Table 2.5 reveals contrasting effects of the variable *house_per* (the proportion of wealth held in housing) on the two stages of the hurdle model, as demonstrated in columns (7) and (8), compared with column (4). These findings suggest nuanced mechanisms influencing household decision-making. Specifically, in the first hurdle, the negative coefficient indicates a crowding-out effect on the probability of participation (extensive margin). This implies that as the proportion of housing expenditure increases, households are less likely to participate in the activity or market under consideration. A plausible explanation is that higher housing costs, being a significant and relatively inflexible component of household budgets, limit financial flexibility, thereby reducing the likelihood of engaging in additional activities. Conversely, in the second hurdle, the positive coefficient reflects a wealth effect on the level of participation (intensive margin). For households that do choose to participate, higher housing expenditure may serve as a proxy for overall wealth or socioeconomic status, enabling greater resource allocation to the activity. This aligns with the notion that wealthier households, despite incurring higher housing costs, may have more disposable income or access to financial resources, facilitating higher levels of engagement. These findings underscore a dual mechanism. On the one hand, higher housing expenditure acts as a constraint during the decision-to-participate stage, reducing participation probabilities (extensive margin). On the other hand, it signals affluence at the participation-intensity stage, promoting greater engagement levels (intensive margin). This bifurcated dynamic provides a deeper understanding of how housing expenditures influence household behavior across different stages of decision-making.

To determine the best empirical model according to my sample, I implement information criteria and likelihood ratio tests for model selection to select which results we should believe with a statistical reason. The information criteria table shows the below Table 2.6 comparing the CDH and the standard Tobit model, the

---

[5]The evidence of variables (*labnum*, *health_bad_ratio*, and *ln_tot_asset*) have conditional variance, or contributing variability to the error-variance function is in Tables 2.10 to 2.12

value of AICs and BICs to the CDH model is lower. Moreover, I run the Likelihood Ratio test between three CDH models, and the result shows *LR chi2(4) = 9024.95;* the p-value is equal to 0, which means when using the CDH model controls the conditional variance specification has a better "Goodness of fit". Therefore, according to the evidence, in the further results report, I would only report the CDH model while controlling the conditional variance error and robust standard error.

Table 2.6: Information Criteria for Model Selection

| Model | Obs | ll(null) | ll(model) | df | AIC | BIC |
|---|---|---|---|---|---|---|
| Tobit_Ori | 35,623 | -15,398.60 | -10,970.10 | 24 | 21,988.20 | 22,191.73 |
| Tobit_Robust | 35,623 | -15,398.60 | -10,970.10 | 24 | 21,988.20 | 22,191.73 |
| CDH_Ori | 35,623 | -12,115.73 | -7,321.70 | 47 | 14,737.39 | 15,135.99 |
| CDH_Robust | 35,623 | -12,115.73 | -7,321.70 | 47 | 14,737.39 | 15,135.99 |
| CDH_Het | 35,623 | -12,115.73 | -3,640.84 | 51 | 7,383.69 | 7,816.20 |

'll(null)' is the log-likelihood of the null model, which only includes the intercept. 'll(model)' is the log-likelihood of the specified model. 'df' is the degrees of freedom, indicating the number of parameters estimated in the model. 'AIC' is the Akaike Information Criterion. 'BIC' is the Bayesian Information Criterion.

I compute the results from the coefficient level to average marginal effect (AME), not only because the coefficients of the *CDH* model do not have a direct interpretation, but also because the marginal effects could be comparable to different models. This is not entirely informative because the probability and the unconditional mean of consumption, depend on both the first-hurdle and second-hurdle regressors.

Due to the distinctive characteristics of each model, as explored in Section 2.4, the CDH model contrasts with the Tobit model by effectively handling zero observations. It distinguishes between households that are constrained from participating and those that opt not to participate. This differentiation is vital for accurately discerning the impact of various factors on both participation and allocation decisions. Notably, the CDH model demonstrates a higher pseudo$R^2$ value (0.699) compared to that of the Tobit model (0.287), suggesting a superior fit to the data. This indicates that the CDH model more effectively captures the variability within the data, resulting in outcomes that are both more reliable and interpretable. Moreover, the outcomes presented in Tables 2.5 and 2.7 illustrate how the CDH model, by accounting for different types of zero observations, identifies distinct influences at both the coefficient and marginal levels, exerting a more substantial impact compared to the Tobit model. The CDH model also exhibits greater robustness to violations of restrictive assumptions such as homoscedasticity and normality in error terms, enhancing its suitability for complex datasets. By capitalising on these advantages, the CDH model provides a more detailed and nuanced insight into household financial decisions than the Tobit model.

In Table 2.7, to the variable of labour risk (*labnum*). Its value is 0.056 significant

at the 1% level, which means when a household has one more stable workforce, the household would allocate more family assets into the financial risk market by 5.6%. Relatively speaking, when a family has a more stable labour force, its ability to hedge risks is better in terms of reducing risk vulnerability. This is relatively consistent with the results of Heaton and Lucas (2000); Betermier et al. (2012); Lin and Lu (2012).

Table 2.7: The Marginal Effects of the Variables of Interest on Dependent Variable

|  | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
|  | Tobit_Robust | | CDH_Het | |
|  | b | se | b | se |
| labnum | 0.013*** | 0.002 | 0.056*** | 0.006 |
| health_bad_ratio | 0.002 | 0.003 | 0.044*** | 0.008 |
| health_insur_ratio | 0.009** | 0.005 | 0.007 | 0.009 |
| house_if | -0.020*** | 0.003 | -0.034*** | 0.006 |
| house_per | -0.118*** | 0.004 | -0.109*** | 0.008 |
| business_if | -0.036*** | 0.003 | -0.043*** | 0.006 |
| business_asset | -0.003*** | 0.001 | -0.003*** | 0.001 |
| Control Variable | YES | | YES | |
| N | 35,623 | | 35,623 | |
| pseudo $R^2$ | 0.287 | | 0.699 | |

Note: Because the reported results table would exceed the page length if it includes control variables; therefore, the full table is reported in the Table 2.17. ***Significant at 1%; **Significant at 5%; *Significant at 10%.

From the perspective of health risks, if a family expects its own health to worsen, the family may invest less in the financial market. Not only does this study not provide empirical evidence of this theoretical implication, but when a family's expectation of its health is poor, this will also lead to an increase in the allocation of its financial risk assets by 4.4%, which is a similar result provides by (Love and Smith, 2010). From the perspective of life cycle theory, when a family holds insurance or life insurance, to a certain extent, it hedges the risk, which also allows the family to readjust the allocation of assets (Baek and DeVaney, 2005). Love and Smith (2010); Goldman and Maestas (2013) believe that health expenditure is exogenous from family asset allocation, and if health expenditure occurs, it will affect the cash flow of investment. Goldman and Maestas (2013) find that individuals who participated in employee health insurance plans held 6% more risky assets than those who did not. Rosen and Wu (2004) believe that health risk and medical expenditure are two related variables that simultaneously affect household financial decision-making.

From the perspective of real estate risk and business risk, due to the dual characteristics of housing, that is, both consumption and investment, it appears not only in the objective function but also in the constraints of the family's maximum

lifetime consumption utility under the given wealth conditions. When real estate forms can be used as collateral, if the proportion of household assets in real estate increases by one unit, the household's allocation of financial risk assets will decrease by 10.9% on average. My results are also consistent with this finding. Cocco (2005) found that the purchase of housing has a crowding-out effect on the purchase of risky assets. From the results of business risk, I find if a family has business ownership and more asset proportion in business, it will lead to less family asset allocation by 4.3% and 0.3%, respectively. These are consistent with Palia et al. (2009), if a household's non-financial income or wealth (such as private business, personal income, real estate) fluctuates less (larger). Then the family is more (smaller) likely to enter the financial risk market and invest a larger (smaller) part of its wealth in the risky market.

## 2.6  Robustness Checks

In line with the same concept as detecting conditional variance, every independent variable is grouped according to the 50 percentile, for group mean testing to explore feasible and potential nonlinear relationships. Similarly, *labnum* and *health_bad_ratio* are statistically unequal at mean.[6]  Therefore, these would raise the following two questions: first, with the current Chinese households' characteristics, would an optimal stable working force number exist? Second, would there exist a point where Chinese households would be more risk-lover? I add the quadratic term of *labnum* and *health_bad_ratio* in the CDH Model, the symbol sign as *labnum_sq* and *health_bad_ratio_sq*. Moreover, in addition to the specific impact of urban or rural household registration on labour risk, housing risk might be another concern in the Chinese context. Because of the different levels of cognition and living environment, urban households generally have more access to risky financial markets, while rural households are more vulnerable to labour risk and house risk. Therefore, I use the variables *rural* interact with *labnum* or *house_if* and *house_per* to create interaction terms, respectively, to check the different effects imposed by the previously mentioned risks in a different identity. The results are as follows:

The full table results of *CDH* model at coefficient level is reported in Tables 2.18 and 2.19. The marginal effect table is shown in Table 2.20. Therefore, we can calculate the optimal stable working force number, which is 2.25, meaning that in China, the optimal stable worker number in a household would be 2 (from 2.25 round-down). The graph evidence is shown in Figure 2.2. When the number of stable workers in a

---

[6]The evidence of variables (*labnum*, *health_bad_ratio*) have statistically unequal at mean in Tables 2.13 and 2.14.

household is fewer than two, there is evidence to suggest that an increased workforce correlates with decreased risk vulnerability for the family. However, if the number of stable workers exceeds two, this trend reversely impacts investment in financially risky assets, leading to a negative effect.

Figure 2.2: The Optimal Stable Workers number at Marginal Effect Level
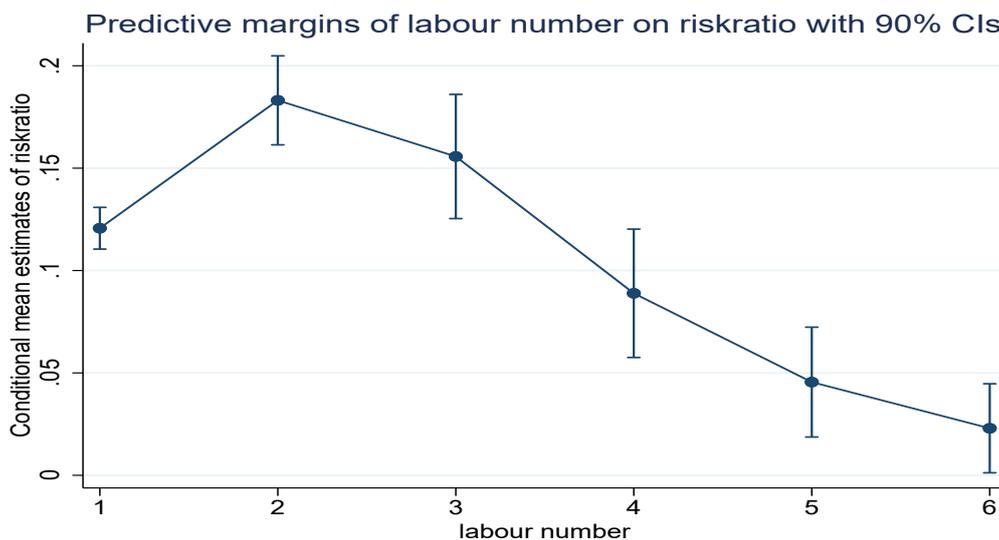


*Note: The x-axis represents the labnum variable, ranging from 1 to 6. The y-axis shows the effects on the conditional mean estimates of riskratio. Each point on the graph indicates the average marginal effect of labnum, with vertical lines representing the 90% confidence intervals.*

The health risk is an interesting part of background risks. As pointed out by Edwards (2008), health shocks may increase or decrease the marginal utility of consumption. In the real world, health shocks may affect the mean and variance of the survival function. In this case, it can be shown that by holding duration constant, an increase in survival risk can actually increase some households' optimal risk shares (Love and Smith, 2010). Therefore, in a specific situation in China, the evidence in Figure 2.3 shows that when a family receives a small amount of health risk shock (the proportion of family members with poor health is about 20%), it is possible to increase the allocation of financial risk assets. However, when the proportion of members with poor health exceeds about 60%, the allocation of financial risk assets will be reduced. The full table results of *CDH* model at coefficient level is reported in Tables 2.21 and 2.22, at marginal effect level is reported in Table 2.23.

In Table 2.8 displays the marginal level result of the interaction term of *rural* and *labnum*. This may have some policy implications. If policymakers want to encourage risky financial asset investment, more attention should be paid to rural areas. But it faces a dilemma in reality. First is that rural residents might have limited financial

Figure 2.3: The marginal plot of the quadratic term of *health_bad_ratio*
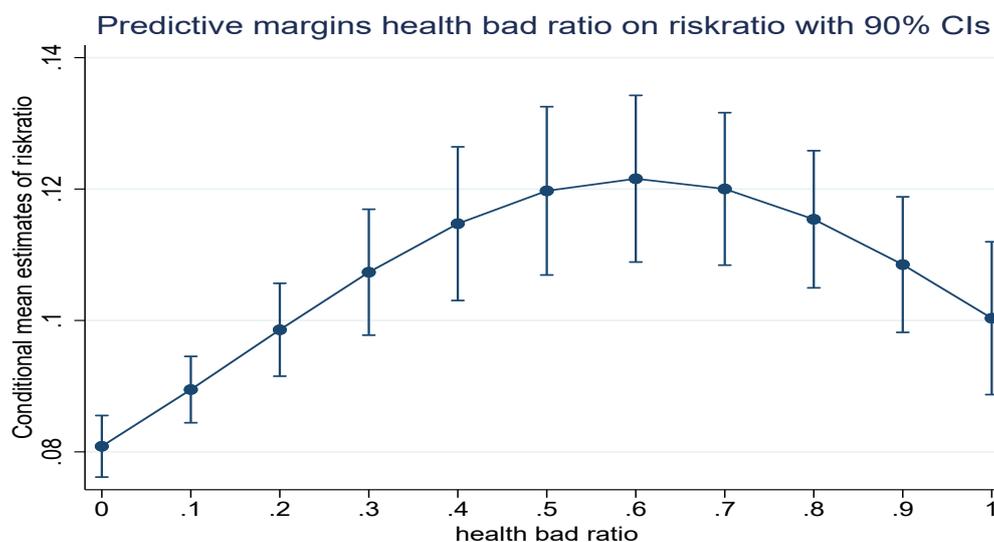


Note: The x-axis represents the health_bad_ratio variable, ranging from 0 to 1. The y-axis shows the effects on the conditional mean estimates of riskratio. Each point on the graph corresponds to the average marginal effect of health_bad_ratio, with vertical lines indicating the 90% confidence intervals.

resources to invest in the capital market. Second is risky asset investment might need abundant knowledge if they want to gain returns. For example, in rural areas, there are some bonuses granted to people living under the absolute poverty line. One qualification to get this bonus is not owning any stock market account. From Table 2.8, we can see the different marginal effects in different regional identities. When the number of stable workers increases by one unit (which is one stable worker), the investment in financial risk assets will increase by 6.4% in urban areas on average while 4.0% in rural areas. The difference is 2.4%, we could relatively say that for urban and rural households, urban households have more chances to allocate family assets to financial risk assets.

Table 2.8: Variables that uniquely identify margins: *rural*

| | dy/dx | std. err. |
|---|---|---|
| labnum | | |
| _at | | |
| 0 | 0.064*** | 0.008 |
| 1 | 0.040*** | 0.007 |

Note: The full table results of *CDH* model at coefficient level is reported in Tables 2.24 and 2.25. ***Significant at 1%; **Significant at 5%; *Significant at 10%.

In addition to the difference in facing house risk when households have either registration on urban or rural, the results are as follows:

From Table 2.9, we could see the impacts of the house are different varying across

Table 2.9: Variables that uniquely identify margins: *rural*

|  | dy/dx | std. err. |
|---|---|---|
| 0.house_if | (base outcome) | |
| 1.house_if | | |
| _at | | |
| 0 | -0.020* | 0.010 |
| 1 | -0.045*** | 0.009 |
| house_per | | |
| _at | | |
| 0 | -0.147*** | 0.012 |
| 1 | -0.062*** | 0.007 |

Note: The full table results of *CDH* model at coefficient level is reported in Tables 2.26 and 2.27. Note: *dy/dx* for factor levels is the discrete change from the base level. ***Significant at 1%; **Significant at 5%; *Significant at 10%.

different identities. When a family owns a house, the urban family would be more likely to allocate family assets to financial risk assets than rural ones, even if the value at the marginal level is negative. However, when the value of the house takes more proportion of the total asset, the family would have less willingness to participate and invest in financial risk assets. To an urban family, it would reduce by 14.7% allocating family assets to financial risk assets, to a rural would reduce by 6.2%. These findings are consistent with the literature, where home ownership tends to lead to a reduced allocation of financial assets to venture capital, and the higher the ratio of home value to wealth, the lower the venture capital (Fratantoni, 1998; Arrondel and Savignac, 2015; Hu et al., 2023). The reason behind the difference between rural and urban households' housing purchasing is straightforward to understand. In China, rural houses are restricted from being transacted and the owners only have the property of using and leasing rights. In this sense, the value of the houses is difficult to evaluate. Situations in urban areas are quite different. Over the past decades, China's urban houses witnessed a great value increase, especially after the implementation of the four-million stimulus policy in 2008 and the housing reconstruction plan in 2015. It encouraged peasants to buy houses in urban areas and entitled these houses to transaction rights. The values of urban houses account for a large part of households' total wealth. This phenomenon is driven by China's two-digit, rural-urban governing system. The background information offers valuable information to understand the reverse coefficients of whether holding a house in rural areas affects participation rates in risky assets. On average, urban citizens are richer than rural citizens. However, rural citizens who benefited from the housing reconstruction policy are richer than those who have not. In this case, in rural areas, citizens who do not have a rural house are more willing to participate in the risk

asset market.

## 2.7 Policy Implications

These findings emphasise the complex interplay between labour, health, housing, and business risks in shaping households' engagement with financial risk assets. The insights gained from analysing Chinese households' financial risk allocation under diverse background risks present valuable implications for policymakers and financial institutions.

To enhance labour market stability, local governments should implement policies that promote job creation in rural areas, such as encouraging local entrepreneurship and investing in infrastructure projects. Strengthen employment protections and develop unemployment insurance schemes to reduce income volatility. Introduce job training initiatives to equip workers with skills aligned to market demands, boosting confidence in future income stability and financial investments.

To mitigate health-related risks, financial institutions should develop comprehensive health insurance schemes that reduce out-of-pocket medical expenses, particularly for low-income and rural households. Government should incentivise regular health check-ups and public health education campaigns to minimise health shocks that could destabilise household finances. Further, by conducting campaigns educating families on how health risks affect financial stability and providing guidance on integrating these considerations into investment decisions.

The government and financial institutions should navigate households balancing housing risk and financial investment decisions. Reassess housing policies to address the over-reliance on real estate as a store of wealth. Encourage the diversification of assets through tax incentives for financial investments or subsidised access to diverse investment products for families heavily invested in real estate.

To reduce the burden of business risk on financial allocations, offer targeted support for small business owners, such as low-interest loans, business insurance, and financial planning services. Providing tailored investment advice to business-owning families can help them balance business-related risks with broader financial market participation.

Lastly, they should enhance financial education and awareness and launch nationwide financial literacy programs that educate families on managing background risks and their impact on financial stability. Delivering tools such as online financial health assessments and personalised risk advisory services can empower households to make informed investment choices. Financial institutions should develop products tailored to the unique risk profiles of different household groups.

## 2.8 Conclusion

The results of this study show that the inclusion of background risks significantly affects household participation and asset allocation in financial markets. Unlike the standard models (binary response model, the multiple linear regression model, Tobit model) commonly adopted in the literature, which often assumes universal participation and homogeneous behaviour across households, the use of Cragg's Double-Hurdle model allows for the differentiation between the decision to participate in the risky asset market and the extent of that participation.

By analysing the impact of Chinese family background risks, which are not easy to avoid or diversify in the allocation of financial risk assets, we can find that if the household has a more stable labour force, the household's risk vulnerability will be relatively reduced, and the household is willing to transfer (invest) more assets to financial risky assets. And this result still holds among urban or rural households, and on average, if a family has two people with stable jobs, labour risk (income risk) is minimised. On average, labour risk is reduced better when a household's labour force comes from urban areas than from rural areas.

When a family self-assesses that the proportion of family members in poor health is higher, the volatility of the family's risk will increase, which will also cause the family to reduce the allocation of financial risk assets. However, when this proportion is around 20%, on average, Chinese households would be more risk-lover, and this will decrease the consumption utility when this proportion is around 60%. For housing risk, at an average level, if a family owns a house, or if a family's house value accounts for a higher proportion of the family's total assets, the overall housing risk will cause the family to reduce the allocation of financial risk assets. This result is consistent between the two samples, identified by the interaction of *rural* and housing risk. If it is a rural family, holding real estate will reduce its allocation of financial risk assets by 4.5%, while urban families will decrease it by 2.0%. If the house value held by a family accounts for 1% more of the total family assets, the financial risk asset allocation can be decreased by 6.2% for rural households and 14.7% for urban households. This may be due to China's current housing-related policies. For business risk, if a household runs a business, its willingness to decrease its financial risk allocation is 4.1%, but if its business assets are worth more, this will reduce its allocation by 0.3%, on average.

This study helps us gain a better understanding of how families allocate their financial risk assets when they encounter background risks. The results of Tobit can be understood as the actual allocation of Chinese households to the financial risk market due to different background risks, while the results of the CDH model can be understood as considering potential households and the actual situation. Therefore,

the difference between the results of the two models is of concern to policymakers and financial institutions. Firstly, policymakers should concentrate on both the economic outcomes and the risk-related consequences for households themselves. They should let people be aware of the potential consequences of ignoring or not well-assessing background risks while making investment decisions. Providing relevant education and reminding people to take a more cautious approach to background risks would benefit households' financial stability. Secondly, as financial institutions offer a wider range of investment products and financial services to Chinese families, they should tailor these offerings according to each family's unique circumstances. This would enhance the participation of families in the financial market, especially those with limited engagement.

Lastly, while the study provides valuable insights into how specific background risks, such as labour income, health, housing, and business, affect household financial risk asset allocation, it has certain constraints. For instance, the data may not fully capture the dynamic nature of risk preferences over time or account for potential interactions between various types of risks. Moreover, the analysis may be limited by the availability of accurate and comprehensive data on household risk-taking behaviors. The use of the CDH model, while suitable for this analysis, might oversimplify the relationship between financial market participation and the extent of that participation, potentially missing nonlinear interactions or dependencies. Moreover, the reliance on data from the 2017 China Household Finance Survey (CHFS) limits the temporal relevance, as it does not account for recent economic developments or shifts in household behaviour. Future research could build on this work by using panel data, adding the time series dimension would be valuable for observing trends over time, providing deeper insights into how household financial behaviours evolve in response to economic changes.

# 2.A   Appendix

## 2.A.1   Additional Tables

Table 2.10: The 50 percentile group Standard Deviation test by *ln_tot_asset*

| Variance ratio test | | | | | | |
|---|---|---|---|---|---|---|
| Group | Obs | Mean | Std. err. | Std. dev. | [95% conf. interval] | |
| 0 | 17,815 | 0.020 | 0.001 | 0.117 | 0.018 | 0.022 |
| 1 | 17,808 | 0.107 | 0.002 | 0.245 | 0.103 | 0.111 |
| Combined | 35,623 | 0.063 | 0.001 | 0.197 | 0.061 | 0.065 |
| ratio = sd(0) / sd(1) | | | | f= 0.228 | | |
| H0: ratio = 1 | | | | Degrees of freedom = 17814, 17807 | | |
| Ha: ratio <1 | Ha: ratio != 1 | | | Ha: ratio >1 | | |
| Pr(F <f) = 0.000 | 2*Pr(F <f) = 0.000 | | | Pr(F >f) = 1.000 | | |

Note: Null Hypothesis (H0): The null hypothesis typically assumes that there is no significant difference in the population standard deviation from a hypothesised value or between two or more groups. If it rejects the null, it suggests a significant difference in the population standard deviation from the hypothesised value between groups.

Table 2.11: The 50 percentile group Standard Deviation test by *labnum*

| Variance ratio test | | | | | | |
|---|---|---|---|---|---|---|
| Group | Obs | Mean | Std. err. | Std. dev. | [95% conf. interval] | |
| 0 | 22,816 | 0.037 | 0.001 | 0.153 | 0.035 | 0.039 |
| 1 | 12,807 | 0.111 | 0.002 | 0.250 | 0.107 | 0.116 |
| Combined | 35,623 | 0.063 | 0.001 | 0.197 | 0.061 | 0.065 |
| ratio = sd(0) / sd(1) | | | | f=0.376 | | |
| H0: ratio = 1 | | | | Degrees of freedom = 22815, 12806 | | |
| Ha: ratio <1 | Ha: ratio != 1 | | | Ha: ratio >1 | | |
| Pr(F <f) = 0.000 | 2*Pr(F <f) = 0.000 | | | Pr(F >f) = 1.000 | | |

Note: Null Hypothesis (H0): The null hypothesis typically assumes that there is no significant difference in the population standard deviation from a hypothesised value or between two or more groups. If it rejects the null, it suggests a significant difference in the population standard deviation from the hypothesised value between groups.

Table 2.12: The 50 percentile group Standard Deviation test by *health_bad_ratio*

| Variance ratio test | | | | | | |
|---|---|---|---|---|---|---|
| Group | Obs | Mean | Std. err. | Std. dev. | [95% conf. interval] | |
| 0 | 21,774 | 0.076 | 0.001 | 0.213 | 0.073 | 0.079 |
| 1 | 13,849 | 0.044 | 0.001 | 0.167 | 0.041 | 0.047 |
| Combined | 35,623 | 0.063 | 0.001 | 0.197 | 0.061 | 0.065 |
| ratio = sd(0) / sd(1) | | | | f=1.621 | | |
| H0: ratio = 1 | | | | Degrees of freedom = 21773, 13848 | | |
| Ha: ratio <1 | Ha: ratio != 1 | | | Ha: ratio >1 | | |
| Pr(F <f) = 1.000 | 2*Pr(F >f) = 0.000 | | | Pr(F >f) = 0.000 | | |

Note: Null Hypothesis (H0): The null hypothesis typically assumes that there is no significant difference in the population standard deviation from a hypothesised value or between two or more groups. If it rejects the null, it suggests a significant difference in the population standard deviation from the hypothesised value between groups.

Table 2.13: The 50 percentile group Mean t-test by *labnum*

Two-sample t test with equal variances

| Group | Obs | Mean | Std. err. | Std. dev. | [95% conf. interval] | |
|---|---|---|---|---|---|---|
| 0 | 22,816 | 0.037 | 0.001 | 0.153 | 0.035 | 0.039 |
| 1 | 12,807 | 0.111 | 0.002 | 0.250 | 0.107 | 0.116 |
| Combined | 35,623 | 0.063 | 0.001 | 0.197 | 0.061 | 0.065 |
| diff | | -0.075 | 0.002 | | -0.079 | -0.071 |

| diff = mean(0) - mean(1) | | t = -34.981 |
|---|---|---|
| H0: diff = 0 | | Degrees of freedom = 35621 |
| Ha: diff <0 | Ha: diff != 0 | Ha: diff >0 |
| Pr(T <t) = 0.000 | Pr(T >t) = 0.000 | Pr(T >t) = 1.000 |

Note: Null Hypothesis (H0): The null hypothesis typically assumes that there is no significant difference between the means being compared. If it rejects the null, it suggests a significant difference between the means being compared, indicating a real effect or difference.

Table 2.14: The 50 percentile group Mean t-test by *health_bad_ratio*

Two-sample t test with equal variances

| Group | Obs | Mean | Std. err. | Std. dev. | [95% conf. interval] | |
|---|---|---|---|---|---|---|
| 0 | 21,774 | 0.076 | 0.001 | 0.213 | 0.073 | 0.079 |
| 1 | 13,849 | 0.044 | 0.001 | 0.167 | 0.041 | 0.047 |
| Combined | 35,623 | 0.063 | 0.001 | 0.197 | 0.061 | 0.065 |
| diff | | 0.032 | 0.002 | | 0.028 | 0.036 |

| diff = mean(0) - mean(1) | | t =15.133 |
|---|---|---|
| H0: diff = 0 | | Degrees of freedom =35621 |
| Ha: diff <0 | Ha: diff != 0 | Ha: diff >0 |
| Pr(T <t) = 1.000 | Pr(T >t) = 0.000 | Pr(T >t) = 0.000 |

Note: Null Hypothesis (H0): The null hypothesis typically assumes that there is no significant difference between the means being compared. If it rejects the null, it suggests a significant difference between the compared means, indicating a real effect or difference.

Table 2.15: The Full Table Results of *Cragg's Double Hurdle* and *Standard Tobit* Model at Coefficient Level

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
|---|---|---|---|---|---|---|---|---|---|
| | CDH_Ori | | Tobit_Ori | CDH_Robust | | Tobit_Robust | | CDH_Het | |
| VARIABLES | riskratio | selection_ll | riskratio | riskratio | selection_ll | riskratio | riskratio | selection_ll | lnsigma |
| labnum | 0.106*** | 0.127*** | 0.071*** | 0.106*** | 0.127*** | 0.071*** | 0.114*** | 0.120*** | 0.168*** |
| | (0.032) | (0.016) | (0.009) | (0.034) | (0.016) | (0.009) | (0.030) | (0.015) | (0.008) |
| health_bad_ratio | 0.031 | -0.008 | -0.010 | 0.031 | -0.008 | -0.010 | 0.020 | 0.112*** | 0.212*** |
| | (0.055) | (0.028) | (0.016) | (0.053) | (0.028) | (0.016) | (0.050) | (0.035) | (0.022) |
| health_insur_ratio | -0.025 | 0.111** | 0.052* | -0.025 | 0.111** | 0.052* | -0.054 | 0.110** | |
| | (0.095) | (0.046) | (0.027) | (0.102) | (0.048) | (0.028) | (0.081) | (0.044) | |
| house_if | -0.071 | -0.206*** | -0.116*** | -0.071 | -0.206*** | -0.116*** | -0.127** | -0.197*** | |
| | (0.059) | (0.030) | (0.018) | (0.059) | (0.031) | (0.018) | (0.052) | (0.030) | |
| house_per | 0.0680 | -1.166*** | -0.666*** | 0.0680 | -1.166*** | -0.666*** | 0.121* | -1.114*** | |
| | (0.078) | (0.037) | (0.023) | (0.084) | (0.039) | (0.022) | (0.069) | (0.039) | |
| business_if | -0.062 | -0.332*** | -0.203*** | -0.062 | -0.332*** | -0.203*** | -0.082 | -0.327*** | |
| | (0.061) | (0.031) | (0.018) | (0.063) | (0.033) | (0.020) | (0.052) | (0.030) | |
| business_asset | 0.001 | -0.027*** | -0.016*** | 0.001 | -0.027*** | -0.016** | -0.001 | -0.025*** | |
| | (0.006) | (0.004) | (0.002) | (0.006) | (0.010) | (0.007) | (0.004) | (0.003) | |
| ln_tot_asset | 0.016 | 0.399*** | 0.229*** | 0.016 | 0.399*** | 0.229*** | 0.032** | 0.405*** | -0.056*** |
| | (0.017) | (0.008) | (0.005) | (0.019) | (0.009) | (0.005) | (0.016) | (0.012) | (0.005) |
| edu_high | -0.075 | 0.007 | 0.010 | -0.075 | 0.007 | 0.010 | -0.054 | 0.013 | |
| | (0.068) | (0.030) | (0.018) | (0.064) | (0.029) | (0.018) | (0.057) | (0.028) | |
| edu_college | -0.016 | 0.235*** | 0.150*** | -0.016 | 0.235*** | 0.150*** | -0.015 | 0.221*** | |
| | (0.049) | (0.024) | (0.014) | (0.049) | (0.024) | (0.014) | (0.043) | (0.024) | |
| edu_graduate | 0.084 | 0.367*** | 0.177*** | 0.084 | 0.367*** | 0.177*** | 0.102 | 0.347*** | |
| | (0.096) | (0.064) | (0.035) | (0.088) | (0.066) | (0.033) | (0.087) | (0.061) | |
| attitu_good | 0.091 | 0.008 | 0.013 | 0.091 | 0.008 | 0.013 | 0.072 | 0.011 | |
| | (0.083) | (0.048) | (0.027) | (0.083) | (0.051) | (0.028) | (0.073) | (0.046) | |

Note: ***Significant at 1%; **Significant at 5%; *Significant at 10%.

Table 2.16: Continued Table 2.15

| VARIABLES | (1) CDH_Ori riskratio | (2) selection_ll | (3) Tobit_Ori riskratio | (4) CDH_Robust riskratio | (5) selection_ll | (6) Tobit_Robust riskratio | (7) riskratio | (8) CDH_Het selection_ll | (9) lnsigma |
|---|---|---|---|---|---|---|---|---|---|
| attitu_ok | 0.056 | -0.007 | 0.004 | 0.06 | -0.007 | 0.004 | 0.052 | -0.012 | |
| | (0.070) | (0.038) | (0.022) | (0.071) | (0.039) | (0.022) | (0.062) | (0.036) | |
| attitu_low | -0.044 | -0.250*** | -0.130*** | -0.044 | -0.250*** | -0.130*** | -0.051 | -0.239*** | |
| | (0.063) | (0.030) | (0.017) | (0.068) | (0.030) | (0.018) | (0.055) | (0.028) | |
| knowledge-good | 0.148* | 0.522*** | 0.276*** | 0.148* | 0.522*** | 0.276*** | 0.071 | 0.496*** | |
| | (0.082) | (0.049) | (0.027) | (0.081) | (0.050) | (0.027) | (0.071) | (0.047) | |
| knowledge-ok | 0.094 | 0.399*** | 0.213*** | 0.094 | 0.399*** | 0.213*** | 0.069 | 0.378*** | |
| | (0.070) | (0.039) | (0.022) | (0.070) | (0.040) | (0.023) | (0.062) | (0.037) | |
| married | 0.079 | 0.113*** | 0.076*** | 0.079 | 0.113*** | 0.076*** | 0.122** | 0.115*** | |
| | (0.063) | (0.032) | (0.018) | (0.064) | (0.033) | (0.019) | (0.055) | (0.030) | |
| age-sq | -0.003 | 0.023*** | 0.012*** | -0.003 | 0.023*** | 0.012*** | -0.009 | 0.022*** | |
| | (0.009) | (0.004) | (0.002) | (0.008) | (0.004) | (0.002) | (0.007) | (0.004) | |
| age | 0.019** | -0.041*** | -0.020*** | 0.019** | -0.041*** | -0.020*** | 0.024*** | -0.039*** | |
| | (0.009) | (0.004) | (0.002) | (0.009) | (0.004) | (0.002) | (0.007) | (0.004) | |
| rural | -0.308*** | -0.478*** | -0.311*** | -0.308*** | -0.478*** | -0.311*** | -0.305*** | -0.464*** | |
| | (0.056) | (0.024) | (0.015) | (0.061) | (0.023) | (0.014) | (0.049) | (0.024) | |
| number | -0.096*** | -0.085*** | -0.051*** | -0.096*** | -0.085*** | -0.051*** | -0.103*** | -0.081*** | |
| | (0.020) | (0.009) | (0.005) | (0.021) | (0.009) | (0.006) | (0.018) | (0.009) | |
| ln_ave_income | -0.001 | 0.008*** | 0.004* | -0.001 | 0.008*** | 0.004** | -0.004 | 0.008*** | |
| | (0.006) | (0.003) | (0.002) | (0.006) | (0.003) | (0.002) | (0.005) | (0.003) | |
| Constant | -1.946*** | 1.706*** | 0.840*** | -1.946*** | 1.706*** | 0.840*** | -2.005*** | 1.594*** | |
| | (0.220) | (0.107) | (0.063) | (0.233) | (0.110) | (0.063) | (0.192) | (0.107) | |
| N | 35,623 | 35,623 | 35,623 | 35,623 | 35,623 | 35,623 | 35,623 | 35,623 | 35,623 |
| pseudo $R^2$ | 0.394 | | 0.287 | 0.394 | | 0.287 | 0.699 | | |

Note: The scale of the quadratic term of Age, due to visibility and readability, is divided by 100. ***Significant at 1%; **Significant at 5%; *Significant at 10%.

Table 2.17: The Marginal Effects of the Variables of Interest on Dependent Variable

| | (1) | (2) |
|---|---|---|
| | CDH_Het | |
| | riskratio | |
| | b | se |
| labnum | 0.056*** | 0.006 |
| health_bad_ratio | 0.044*** | 0.008 |
| health_insur_ratio | 0.007 | 0.009 |
| house_if | -0.034*** | 0.006 |
| house_per | -0.109*** | 0.008 |
| business_if | -0.043*** | 0.006 |
| business_asset | -0.003*** | 0.001 |
| ln_tot_asset | 0.035*** | 0.002 |
| edu_high | -0.004 | 0.006 |
| edu_college | 0.022*** | 0.005 |
| edu_graduate | 0.047*** | 0.011 |
| attitu_good | 0.008 | 0.009 |
| attitu_ok | 0.004 | 0.007 |
| attitu_low | -0.031*** | 0.006 |
| knowledge_good | 0.061*** | 0.009 |
| knowledge_ok | 0.048*** | 0.007 |
| married | 0.024*** | 0.006 |
| age_sq | 0.002* | 0.001 |
| age | -0.002** | 0.001 |
| rural | -0.080*** | 0.006 |
| number | -0.019*** | 0.002 |
| ln_ave_income | 0.001* | 0.001 |
| N | 35,623 | |
| pseudo $R^2$ | 0.699 | |

Note: The scale of the quadratic term of Age, due to visibility and readability, is divided by 100. ***Significant at 1%; **Significant at 5%; *Significant at 10%.

Table 2.18: The Full Table Result of Optimal Stable Working Force Number in Chinese Households

| VARIABLES | (1) riskratio | (2) selection_ll | (3) lnsigma |
|---|---|---|---|
| labnum | 0.193*** | 0.194*** | 0.340*** |
| | (0.055) | (0.032) | (0.021) |
| c.labnum#c.labnum | -0.038* | -0.035*** | -0.085*** |
| | (0.020) | (0.012) | (0.009) |
| health_bad_ratio | 0.023 | 0.123*** | 0.189*** |
| | (0.051) | (0.035) | (0.022) |
| health_insur_ratio | -0.041 | 0.106** | |
| | (0.083) | (0.044) | |
| house_if | -0.118** | -0.186*** | |
| | (0.053) | (0.030) | |
| house_per | 0.122* | -1.093*** | |
| | (0.070) | (0.039) | |
| business_if | -0.071 | -0.312*** | |
| | (0.052) | (0.030) | |
| ln_tot_asset | 0.032** | 0.392*** | -0.055*** |
| | (0.016) | (0.012) | (0.005) |
| business_asset | -0.001 | -0.025*** | |
| | (0.004) | (0.003) | |
| edu_high | -0.055 | 0.020 | |
| | (0.058) | (0.028) | |
| edu_college | -0.016 | 0.217*** | |
| | (0.043) | (0.023) | |
| edu_graduate | 0.109 | 0.305*** | |
| | (0.088) | (0.062) | |

Note: ***Significant at 1%; **Significant at 5%; *Significant at 10%.

Table 2.19: Continued table 2.18

| VARIABLES | (1) riskratio | (2) selection_ll | (3) lnsigma |
|---|---|---|---|
| attitu_good | 0.080 | 0.004 | |
| | (0.075) | (0.046) | |
| attitu_ok | 0.055 | -0.017 | |
| | (0.063) | (0.036) | |
| attitu_low | -0.044 | -0.236*** | |
| | (0.056) | (0.028) | |
| knowledge_good | 0.064 | 0.488*** | |
| | (0.072) | (0.047) | |
| knowledge_ok | 0.069 | 0.376*** | |
| | (0.063) | (0.037) | |
| married | 0.111** | 0.111*** | |
| | (0.056) | (0.030) | |
| age_sq | -0.006 | 0.022*** | |
| | (0.007) | (0.004) | |
| age | 0.022*** | -0.039*** | |
| | (0.007) | (0.004) | |
| rural | -0.306*** | -0.457*** | |
| | (0.050) | (0.024) | |
| number | -0.104*** | -0.068*** | |
| | (0.018) | (0.009) | |
| ln_ave_income | -0.341 | 2.396*** | |
| | (0.561) | (0.473) | |
| Constant | -1.999*** | 1.498*** | |
| | (0.196) | (0.107) | |
| $N$ | 35,623 | 35,623 | 35,623 |
| pseudo $R^2$ | 0.702 | | |

Note: The scale of the quadratic term of Age, due to visibility and readability, is divided by 100. ***Significant at 1%; **Significant at 5%; *Significant at 10%.

Table 2.20: Variables that uniquely identify margins: *labnum*

| | dy/dx | std. err. |
|---|---|---|
| labnum | | |
| _at | | |
| 1 | 0.081*** | 0.009 |
| 2 | 0.023* | 0.013 |
| 3 | -0.065*** | 0.013 |
| 4 | -0.057*** | 0.008 |
| 5 | -0.031*** | 0.005 |
| 6 | -0.016*** | 0.004 |

Note: The full table results of *CDH* model at coefficient level is reported in Tables 2.18 and 2.19. ***Significant at 1%; **Significant at 5%; *Significant at 10%.

Table 2.21: The Full Table Result of the Quadratic term of *health_bad_ratio*

| VARIABLES | (1) riskratio | (2) selection_ll | (3) lnsigma |
|---|---|---|---|
| labnum | 0.106*** | 0.106*** | 0.153*** |
|  | (0.027) | (0.014) | (0.008) |
| health_bad_ratio | -0.131 | 0.265** | 0.814*** |
|  | (0.182) | (0.124) | (0.084) |
| c.health_bad_ratio# c.health_bad_ratio | 0.171 | -0.153 | -0.702*** |
|  | (0.189) | (0.129) | (0.093) |
| health_insur_ratio | -0.035 | 0.101** |  |
|  | (0.082) | (0.043) |  |
| house_if | -0.120** | -0.183*** |  |
|  | (0.052) | (0.029) |  |
| house_per | 0.107 | -1.069*** |  |
|  | (0.069) | (0.039) |  |
| business_if | -0.081 | -0.308*** |  |
|  | (0.052) | (0.030) |  |
| business_asset | -0.000 | -0.024*** |  |
|  | (0.004) | (0.003) |  |
| ln_tot_asset | 0.035** | 0.384*** | -0.057*** |
|  | (0.016) | (0.012) | (0.005) |
| edu_high | -0.059 | 0.020 |  |
|  | (0.057) | (0.0280) |  |
| edu_college | -0.016 | 0.218*** |  |
|  | (0.043) | (0.023) |  |
| edu_graduate | 0.097 | 0.303*** |  |
|  | (0.087) | (0.061) |  |

Note: ***Significant at 1%; **Significant at 5%; *Significant at 10%.

Table 2.22: Continued Table 2.21

| VARIABLES | (1)<br>riskratio | (2)<br>selection_ll | (3)<br>lnsigma |
|---|---|---|---|
| attitu_good | 0.058 | 0.008 | |
| | (0.073) | (0.046) | |
| attitu_ok | 0.050 | -0.017 | |
| | (0.062) | (0.036) | |
| attitu_low | -0.048 | -0.233*** | |
| | (0.055) | (0.028) | |
| knowledge_good | 0.083 | 0.481*** | |
| | (0.072) | (0.047) | |
| knowledge_ok | 0.065 | 0.370*** | |
| | (0.062) | (0.037) | |
| married | 0.119** | 0.107*** | |
| | (0.055) | (0.030) | |
| age_sq | -0.010 | 0.021*** | |
| | (0.007) | (0.004) | |
| age | 0.025*** | -0.037*** | |
| | (0.007) | (0.004) | |
| rural | -0.289*** | -0.454*** | |
| | (0.050) | (0.024) | |
| number | -0.103*** | -0.065*** | |
| | (0.019) | (0.009) | |
| ln_ave_income | -0.179 | 2.567*** | |
| | (0.542) | (0.466) | |
| Constant | -2.056*** | 1.466*** | |
| | (0.197) | (0.108) | |
| $N$ | 35,623 | 35,623 | 35,623 |
| pseudo $R^2$ | 0.702 | | |

Note: The scale of the quadratic term of Age, due to visibility and readability, is divided by 100.
***Significant at 1%; **Significant at 5%; *Significant at 10%.

Table 2.23: Variables that uniquely identify margins: *health_bad_ratio*

| | dy/dx | std. err. |
|---|---|---|
| health_bad_ratio | | |
| _at | | |
| 0.0 | 0.088*** | 0.019 |
| 0.1 | 0.099*** | 0.021 |
| 0.2 | 0.104*** | 0.022 |
| 0.3 | 0.097*** | 0.020 |
| 0.4 | 0.077*** | 0.016 |
| 0.5 | 0.044*** | 0.012 |
| 0.6 | 0.003 | 0.016 |
| 0.7 | -0.037 | 0.024 |
| 0.8 | -0.068** | 0.029 |
| 0.9 | -0.086*** | 0.029 |
| 1.0 | -0.092*** | 0.027 |

Note: The full table results of *CDH* model at coefficient level is reported in Tables 2.21 and 2.22. ***Significant at 1%; **Significant at 5%; *Significant at 10%.

Table 2.24: The Full Table of the Result of having Interaction term of *rural* and *labnum*

| | (1) | (2) | (3) |
|---|---|---|---|
| VARIABLES | riskratio | selection_ll | lnsigma |
| 1.rural | -0.360*** | -0.521*** | |
| | (0.062) | (0.031) | |
| labnum | 0.088*** | 0.084*** | 0.168*** |
| | (0.029) | (0.017) | (0.008) |
| 1.rural#c.labnum | 0.081 | 0.080*** | |
| | (0.056) | (0.025) | |
| health_bad_ratio | 0.021 | 0.118*** | 0.212*** |
| | (0.050) | (0.035) | (0.022) |
| health_insur_ratio | -0.051 | 0.112** | |
| | (0.081) | (0.044) | |
| house_if | -0.122** | -0.182*** | |
| | (0.053) | (0.030) | |
| house_per | 0.118* | -1.092*** | |
| | (0.069) | (0.039) | |
| business_if | -0.075 | -0.311*** | |
| | (0.051) | (0.030) | |
| business_asset | -0.001 | -0.025*** | |
| | (0.004) | (0.003) | |
| ln_tot_asset | 0.033** | 0.392*** | -0.056*** |
| | (0.016) | (0.012) | (0.005) |

Note: ***Significant at 1%; **Significant at 5%; *Significant at 10%.

Table 2.25: Continued Table 2.24

| | (1) | (2) | (3) |
|---|---|---|---|
| VARIABLES | riskratio | selection_ll | lnsigma |
| edu_high | -0.057 | 0.014 | |
| | (0.057) | (0.028) | |
| edu_college | -0.019 | 0.217*** | |
| | (0.043) | (0.023) | |
| edu_graduate | 0.104 | 0.308*** | |
| | (0.087) | (0.062) | |
| attitu_good | 0.073 | 0.005 | |
| | (0.073) | (0.046) | |
| attitu_ok | 0.053 | -0.017 | |
| | (0.062) | (0.036) | |
| attitu_low | -0.053 | -0.238*** | |
| | (0.055) | (0.028) | |
| knowledge_good | 0.068 | 0.487*** | |
| | (0.071) | (0.047) | |
| knowledge_ok | 0.070 | 0.376*** | |
| | (0.062) | (0.037) | |
| married | 0.122** | 0.113*** | |
| | (0.055) | (0.030) | |
| age_sq | -0.009 | 0.020*** | |
| | (0.007) | (0.004) | |
| age | 0.024*** | -0.037*** | |
| | (0.007) | (0.004) | |
| number | -0.104*** | -0.068*** | |
| | (0.018) | (0.009) | |
| ln_ave_income | -0.133 | 2.680*** | |
| | (0.550) | (0.475) | |
| Constant | -2.006*** | 1.510*** | |
| | (0.194) | (0.108) | |
| N | 35,623 | 35,623 | 35,623 |
| pseudo $R^2$ | 0.700 | | |

Note: The scale of the quadratic term of Age, due to visibility and readability, is divided by 100.
***Significant at 1%; **Significant at 5%; *Significant at 10%.

Table 2.26: The Difference on having Rural Identity on Housing Risk

| VARIABLES | (1)<br>riskratio | (2)<br>selection_ll | (3)<br>lnsigma |
|---|---|---|---|
| labnum | 0.105*** | 0.106*** | 0.168*** |
| | (0.027) | (0.014) | (0.008) |
| health_bad_ratio | 0.022 | 0.124*** | 0.211*** |
| | (0.050) | (0.035) | (0.022) |
| health_insur_ratio | -0.052 | 0.115*** | |
| | (0.081) | (0.044) | |
| 1.rural | -0.237** | -0.269*** | |
| | (0.097) | (0.052) | |
| 1.house_if | -0.114* | -0.032 | |
| | (0.062) | (0.037) | |
| 1.rural#1.house_if | -0.042 | -0.397*** | |
| | (0.107) | (0.058) | |
| house_per | 0.135* | -1.183*** | |
| | (0.077) | (0.046) | |
| 1.rural#c.house_per | -0.071 | 0.210*** | |
| | (0.129) | (0.061) | |
| business_if | -0.077 | -0.320*** | |
| | (0.051) | (0.030) | |
| business_asset | -0.001 | -0.025*** | |
| | (0.004) | (0.003) | |
| ln_tot_asset | 0.033** | 0.395*** | -0.056*** |
| | (0.016) | (0.012) | (0.005) |
| edu_high | -0.055 | 0.019 | |
| | (0.057) | (0.028) | |
| edu_college | -0.018 | 0.222*** | |
| | (0.043) | (0.023) | |
| edu_graduate | 0.105 | 0.314*** | |
| | (0.087) | (0.062) | |

Note: ***Significant at 1%; **Significant at 5%; *Significant at 10%.

Table 2.27: Continued Table 2.26

| VARIABLES | (1) riskratio | (2) selection_ll | (3) lnsigma |
|---|---|---|---|
| attitu_good | 0.070 | 0.002 | |
| | (0.073) | (0.046) | |
| attitu_ok | 0.050 | -0.026 | |
| | (0.062) | (0.036) | |
| attitu_low | -0.053 | -0.241*** | |
| | (0.055) | (0.028) | |
| knowledge_good | 0.073 | 0.489*** | |
| | (0.071) | (0.047) | |
| knowledge_ok | 0.071 | 0.376*** | |
| | (0.062) | (0.037) | |
| married | 0.121** | 0.105*** | |
| | (0.055) | (0.030) | |
| age_sq | -0.008 | 0.020*** | |
| | (0.007) | (0.004) | |
| age | 0.023*** | -0.036*** | |
| | (0.007) | (0.004) | |
| number | -0.104*** | -0.064*** | |
| | (0.018) | (0.009) | |
| ln_ave_income | -0.195 | 2.543*** | |
| | (0.549) | (0.472) | |
| Constant | -2.027*** | 1.377*** | |
| | (0.197) | (0.110) | |
| $N$ | 35,623 | 35,623 | 35,623 |
| pseudo $R^2$ | 0.701 | | |

Note: The scale of the quadratic term of Age, due to visibility and readability, is divided by 100. ***Significant at 1%; **Significant at 5%; *Significant at 10%.

## 2.A.2 Additional Equations

## 2.A.3 The Marginal Effect Derivation Process of the CDH Model

$$\text{risk\_if}_i^* = \boldsymbol{\alpha} \boldsymbol{X'_i} + \gamma_{1i} \tag{2.12}$$

$$\text{riskratio}_i^{**} = \boldsymbol{\beta} \boldsymbol{X'_i} + \gamma_{2i} \tag{2.13}$$

$$riskratio_i = \begin{cases} riskratio_i^* & \text{,if } riskratio_i^{**} > 0 \,\&\, \text{risk\_if}_i^* > 0 \\ 0 & \text{,otherwise} \end{cases}$$

$$\begin{pmatrix} \gamma_{1i} \\ \gamma_{2i} \end{pmatrix} \sim \boldsymbol{N} \left[ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & \sigma^2 \end{pmatrix} \right]$$

where $\gamma_{1i} \sim N(0,1)$, $\gamma_{2i} \sim N(0,\sigma^2)$ are independent random error terms. Taking Equation (2.3), Equation (2.5) from Section 2.4, and the normality assumption of the error terms, the probability of participation (First Hurdle) and consumption (Second Hurdle) are respectively Equation (2.12) and Equation (2.13):

$$\text{P}\left(\text{risk\_if}_i^* > 0\right) = \Phi\left(x_i \alpha\right) \tag{2.14}$$

$$\begin{aligned} \text{P}\left(\text{riskratio}_i^* >\right) &= \text{P}\left(\left(\text{risk\_if}_i^*, riskratio_i^{**} > 0\right)\right) \\ &= \Phi\left(x_i \alpha\right) \Phi\left(\frac{x_i \beta}{\sigma}\right) \end{aligned} \tag{2.15}$$

In *CDH* model, this is more commonly known as the unconditional expectation (or unconditional mean) of $riskratio_i$ and is written as $\text{E}\left[riskratio_i \mid \text{X}\right]$. The unconditional expectation can be decomposed into two parts, the conditional expectation, $[riskratio_i \mid \text{x}, riskratio_i > 0]$, which is the expected value of $riskratio_i$ for values of the explanatory variables $x$, conditional of $(riskratio_i > 0)$ and the probability of a positive value of $riskratio_i$ for values of the explanatory variables, $x$, $\text{P}\left[riskratio_i > 0 \mid x\right]$. In the CDH model, the probability of participation and the level of investing conditional on participation (Yen and Su, 1995; Mutlu and Gracia, 2006):

$$\text{E}\left[riskratio_i \mid riskratio_i > 0, x\right] = x_i \beta + \sigma \left[\frac{\phi\left(\frac{x_i \beta}{\sigma}\right)}{\Phi\left(\frac{x_i \beta}{\sigma}\right)}\right] \tag{2.16}$$

Thus, the unconditional mean $riskratio_i$ is:

$$E\left(\text{riskratio}_i >| \; x\right) = P\left(\text{riskratio}_i > 0\right) E\left(\text{riskratio}_i \mid \text{riskratio}_i > 0\right)$$

$$= \Phi\left(x_i \alpha\right) \Phi\left(\frac{x_i \beta}{\sigma}\right) \left\{ x_i \beta + \sigma \left[ \frac{\phi\left(\frac{x_i \beta}{\sigma}\right)}{\Phi\left(\frac{x_i \beta}{\sigma}\right)} \right] \right\} \qquad (2.17)$$

Marginal effects can be calculated by differentiating each of the above equations with respect to each explanatory variable. These equations are given by:

$$\frac{\partial P\left[risk\_if_i > 0 \mid x\right]}{\partial x_j} = \alpha_j \phi\left(x_i \alpha\right) \qquad (2.18)$$

$$\frac{\partial P\left[riskratio_i > 0 \mid x\right]}{\partial x_j} = \alpha_j \phi\left(x_i \alpha\right) \Phi\left(\frac{x_i \beta}{\sigma}\right) + \Phi(x_i \alpha)\phi\left(\frac{x_i \beta}{\sigma}\right)$$
$$\frac{1}{\sigma}\left[\beta_j - \left(\frac{x_i \beta}{\sigma}\right) \frac{\partial \sigma}{\partial x_j}\right] \qquad (2.19)$$

$$\frac{\partial E\left[riskratio_i \mid riskratio_i > 0, x\right]}{\partial x_j} = \beta_j + \left[\frac{\phi\left(\frac{x_i \beta}{\sigma}\right)}{\Phi\left(\frac{x_i \beta}{\sigma}\right)}\right] \frac{\partial \sigma}{x_j} - \left[\frac{\phi\left(\frac{x_i \beta}{\sigma}\right)}{\Phi\left(\frac{x_i \beta}{\sigma}\right)}\right]$$
$$\left[\beta_j - \left(\frac{x_i \beta}{\sigma}\right) \frac{\partial \sigma}{\partial x_j}\right] \left[\frac{x_i \beta}{\sigma} + \frac{\phi\left(\frac{x_i \beta}{\sigma}\right)}{\Phi\left(\frac{x_i \beta}{\sigma}\right)}\right] \qquad (2.20)$$

The marginal effect for the unconditional level can be derived by applying the product rule of differentiation to the above equations:

$$\frac{\partial E\left[riskratio_i > 0 \mid x\right]}{\partial x_j} = \frac{P[riskratio_i > 0 \mid x]}{\partial x_j} E[riskratio_i \mid riskratio_i > 0, x]$$
$$+ P(riskratio_i > 0 \mid x)\frac{\partial E[riskratio_i \mid riskratio_i > 0, x]}{\partial x_j}$$
$$(2.21)$$

where is the marginal effect of the conditional expectation equals the marginal effect of the probability of positive value times the conditional expectation plus the marginal effect of the conditional expectation times the probability of a positive value.

The coefficients' significance level might differ from that in the marginal effects. This might be related to the distribution of the independent variables. Since the marginal effect is usually referred as the average marginal effect at different points of the independent variable, the marginal effect is then thus a non-linear function of the coefficients.

# Chapter 3

# Investigating the Simultaneous Relationship between Migrants' Mobility and Life Satisfaction

Research on migration mobility initially emphasised employment opportunities, better jobs, and higher earnings for individuals. However, emerging studies propose that migration's impact on life satisfaction, beyond utilitarian considerations, should also be considered. This study presents empirical evidence on the reciprocal relationship between life satisfaction and individual migration in China. Using data from the China Family Panel Studies (CFPS) 2012–2020 panel, it examines how changes in life satisfaction may drive mobility decisions and how mobility decisions, in turn, impact life satisfaction. When life satisfaction is low, individuals are 29 p.p. more likely to make mobility decisions than when life satisfaction is very low. To a higher life satisfaction level, individuals are not likely to move. The evidence shows that, compared with individuals who did not move, migration decision has a positive association with life satisfaction. These findings highlight the pivotal role of individual life satisfaction, serving not only as a significant determinant of migration decisions but also migrants making decisions to improve life satisfaction.

## 3.1   Introduction

Over the past decades, China has undergone substantial human migration from rural to urban areas. Extensive literature has consistently identified economic factors, such as low wages and unemployment, as predominant catalysts for labour migration (Hu, 2012; Gu, 2014; Zheng and Yang, 2016; Tan et al., 2017). However, recent years have witnessed a discernible shift in migration patterns from eastern cities to central and western regions in China (Chuankai and Yuemin, 2015; Wang, 2016b). According to data from the National Bureau of Statistics, the count of migrant workers moving from eastern China to central and western regions surged from 15.7 million in 2010 to 28.2 million in 2019, indicative of the escalating migration trend from the east to the central and western sectors.

If we assume that higher wage rates and better employment opportunities are the primary drivers of migration from underdeveloped regions to major cities, the decision to migrate from developed eastern areas to central or western cities becomes a particularly intriguing phenomenon. Rural-to-urban migration in China is associated with heightened perceptions of relative deprivation, resulting in lower levels of happiness and diminished emotional well-being compared to non-migrants (Jin, 2016). Post-2009, as China transitioned from a lower-middle-income to an upper-middle-income country, income-related disparities in happiness intensified. Yang et al. (2019) reveal that an increase in income inequality intensifies disparities in happiness satisfaction states, while enhancements in income and education levels tend to mitigate these disparities. This observation aligns with the theory of relative deprivation, which suggests that migration decisions are influenced not only by wage differentials and improved educational and employment opportunities but also by life (dis)satisfaction, which can serve as a significant determinant of migration intentions (Stark and Yitzhaki, 1988; Otrachshenko and Popova, 2012). However, the theory of return migration posits that migrants decide to return after accumulating sufficient savings and attaining their objectives(Dustmann, 1996). Analogous to the theory of how (dis)satisfaction influences migration decisions (Stark and Yitzhaki, 1988; Otrachshenko and Popova, 2012) and in alignment with the theory of keep moving migration, individuals make these decisions when the expected net utility is positive. So, particularly in the context of the subjective element of people's well-being, the question of whether immigrants exhibit greater satisfaction with living conditions or increased happiness compared to those who remain in their place of origin is a nuanced inquiry.

In this chapter, I explore whether life satisfaction significantly influences mobility decisions and, conversely, whether mobility enhances life satisfaction, focusing on a

case study in China. Additionally, this analysis considers whether life satisfaction drives return migration and if life satisfaction among migrants improves upon leaving major urban centers. Moreover, the study examines the varied impacts of educational levels on migration decisions as moderated by life satisfaction, employing the construction of interaction terms. To my knowledge, this study is the first to use panel data from the China Family Panel Studies (CFPS) from 2012 to 2020 to present empirical evidence on the bidirectional relationship between life satisfaction and individual migration within China. The analysis seeks to understand how changes in life satisfaction may precipitate mobility decisions and how these decisions, in turn, influence life satisfaction. Given the nature of the data, migration decisions are treated as a binary variable, and life satisfaction is considered a categorical variable, necessitating the use of panel Probit and ordered Probit estimations, respectively. To address potential endogeneity issues, instrumental variables are employed. The chapter will also revisit the motivations behind return migration and assess whether migrants' life satisfaction increases when they move away from large urban areas. Additionally, it will delve into how educational attainment diversely affects migration choices, mediated by life satisfaction, through detailed interaction terms. The analysis reveals a 6.3 p.p. decrease in the willingness to migrate among satisfied individuals. Mobility in search of enhanced life satisfaction is associated with subsequent positivity in life satisfaction. These findings highlight the pivotal role of individual life satisfaction, serving not only as a significant determinant of migration decisions but also as a mediator linking individual socioeconomic variables and macroeconomic conditions.

The extensive literature on this topic offers numerous examples, debates, and detailed analyses, highlighting the complex factors influencing subjective well-being in migrant behaviour (Wright, 2011; Hendriks and Bartram, 2019). Bartram (2013) investigates the relationship between happiness, as an indicator of subjective well-being, and emigration from 14 Eastern European countries to Western Europe. Bartram's study indicates that Eastern European emigrants generally experience increased happiness, attributed more to pre-existing higher happiness levels than purely economic motives. While emigrants tend to report higher happiness, notable variations exist across countries, with Polish emigrants showing comparable happiness levels to non-emigrants. Papastergiou et al. (2023) examines the link between subjective well-being and migration using panel data from the British Household Panel Survey (BHPS) and the UK Household Longitudinal Study (UKHLS) from 2004-2016, finding no significant impact on natives' well-being. The causal relationship between migration mobility and well-being or life satisfaction has drawn considerable attention from researchers across disciplines and has become

crucial for policy considerations. Some empirical studies investigate this relationship but often neglect the bidirectional effects of migration mobility on life satisfaction. Our research addresses this gap. Furthermore, the literature on life satisfaction and migration decisions in the context of internal migration, particularly in China, remains limited. Our research aims to provide robust evidence in this area.

The content of this chapter is arranged as follows: Section 3.2 reviews the relevant literature; Section 3.3 describes the database, main variable relations, Section 3.4 the empirical models; Section 3.5 reports the data analysis; Section 3.8 Conclusion.

## 3.2 Literature Review

This literature review will cover three main parts of the topic of life satisfaction and migration.

### 3.2.1 The Concept of Life Satisfaction

In the realm of subjective well-being research, Veenhoven (1995); Selezneva (2011); Bartolini (2011) propose that happiness signifies the extent to which an individual appraises their entire life positively, which refers to the level of contentment a person experiences with the life they are leading. They further elaborate that subjective well-being corresponds to what can be termed as "experienced utility," encompassing a reflective assessment of the attained utility. Additionally, life satisfaction is characterised as a discernment regarding objective circumstances and various domains of life. According to an OECD compilation, a life satisfaction indicator measures how people perceive their lives as a whole in terms of overall life satisfaction.

The evaluation of overall life satisfaction encompasses various perspectives, including emotions, expectations, achievements, and more. Different studies have interchangeably referred to life satisfaction as subjective well-being (SWB), indicating their equivalence. Veenhoven (2011) employs the conscious facet of subjective well-being, namely life satisfaction, as a means of quantifying the influence of utility maximization on immigrants' decision-making.

In terms of temporal validity, Michalos and Maurine Kahlke (2010) suggest that unidimensional expressions of life satisfaction exhibit less stability over time compared to multidimensional metrics, which maintain a stronger correlation over extended periods. Furthermore, Diener et al. (2013) and Michalos et al. (2017), while reviewing life satisfaction measures used in research, find that the inclusion of additional social and economic indicators diminishes measurement errors in life

satisfaction indicators. An additional facet to consider involves the types of measures, scales, or scores, and their effectiveness in assessing life satisfaction.

Considering the validity of measurement tools, Diener et al. (1985), Pavot and Diener (1993) and Neto and Fonseca (2016) delve into the validity of the Satisfaction with Life Scale (SWLS).[1] Nevertheless, the SWLS is recommended for its more robust psychometric foundation. Cheung and Lucas (2014) undertake an evaluation of the validity of single-item life satisfaction measures by comparing them to the SWLS.[2] The findings indicate that single-item life satisfaction measures perform comparably to the multi-item SWLS, indicating that social scientists would obtain nearly identical responses to substantive inquiries regardless of the measurement employed.

Within my specific context, I have the opportunity to leverage survey data collected in China between 2012 and 2020 to discern the individuals who opted for migration during this time frame and reported their subjective well-being. Consequently, I employ the term "life satisfaction" to denote the level of satisfaction expressed at a specific instance in time, which aligns with the moment of the interview. In this particular context, life satisfaction adheres to a single-item measurement.

## 3.2.2   The Rationale Behind Employing SWB in Migration

The existing body of literature provides a solid foundation for establishing a connection between subjective well-being and utility (Bartolini, 2007; Dolan et al., 2008; Frey, 2018). Although, utility is inherently subjective, as it is an individual's evaluation of the satisfaction or well-being obtained from objective conditions. Recognising the inherent subjectivity of value, it becomes that human traits, which are greatly influenced by cultural factors, may affect the appraisal of utility. The influence mentioned can be attributed to cross-cultural variances, which can be effectively explained by disparities in cultural values and self-construal (Diener et al., 1999; Cheng et al., 2016).

However, it is crucial to note that acknowledging this subjectivity does not imply

---

[1]The Satisfaction with Life Scale (SWLS) is a comprehensive tool for assessing life satisfaction. It is a short 5-item instrument designed to measure global cognitive judgments of satisfaction with one's life. Scores consist of a raw score (between 5 and 35). Higher scores represent higher life satisfaction. Scorers can be assigned into six well-being categories and interpretative text is provided for each, 30-35 Extremely satisfied; 25–29 Satisfied; 20–24 Slightly satisfied; 15–19 Slightly dissatisfied; 10–14 Dissatisfied; 5–9 Extremely dissatisfied

[2]A single-term life satisfaction measure is a simple question or statement that individuals are asked to respond to in order to assess their overall life satisfaction. It typically involves using a single rating scale or response option to gauge how satisfied an individual feels with their life as a whole. For example, a common single-item measure might be: "On a scale of 1 to 5, how satisfied are you with your life?" Participants would then choose a number on the scale that best represents their level of life satisfaction, with higher numbers indicating higher satisfaction.

that subjective well-being is inherently an unreliable measure of utility, which is true when there is a significant discrepancy between individuals' subjective well-being assessments and their real emotional experiences. Moreover, if this disparity consistently varies across different countries or regions, it can undermine the reliability of SWB as a measure. However, a considerable number of studies have extensively utilised SWB in various cultural contexts, and the proven dependability of SWB as a metric of value within a certain civilisation or even across different nations. (Hadjar and Backes, 2013; Samuel and Hadjar, 2016; Glatz and Eder, 2020). The evidence suggests that SWB is a fairly reliable measure that may be effectively applied in these circumstances, even in different countries.

Another distinctive attribute of subjective well-being, particularly when scrutinising migration decisions, lies in its association with the degree of fulfillment of personal goals and expectations (Diener and Fujita, 1995; Kökönyei et al., 2008; Rachmat et al., 2019). Empirical investigations have substantiated that migration decisions are influenced not only by wage disparities and improved educational and employment prospects but also by levels of life satisfaction or dissatisfaction, which can act as robust determinants of migration intentions(Stark and Yitzhaki, 1988). According to the proposition, an individual's subjective well-being is dependent on the degree to which their aspirations and expectations are fulfilled. The inverse is also true if an individual achieves their expectations, their SWB tends to rise. Migrants often hold certain expectations, whether established before migration or formed after migration, that are later either met or unfulfilled in the new host society. The aforementioned phenomenon inherently influences the subjective well-being of immigrants residing in the host region, hence establishing SWB as a significant measure for evaluating the fulfillment of expectations.

The prior researchers aimed to address this inquiry by employing income, or absolute income, and other objective metrics of utility. This approach tends to align more closely with Neoclassical theory or emphasise a cost-benefit analysis (Constant and Massey, 2002; Bijwaard and Wahba, 2014; Bijwaard, 2015). However, it might be insufficiently clear to ascertain if income was the sole driving motivation behind migration, which implies that considering persistent migration solely in terms of income can lead to potential ambiguities. For instance, some researchers posit that a decline in income may correspondingly lead to an increase in floating migration, thus postulating an intricate relationship between income and migration outcomes(Sjaastad, 1962; Renas and Kumar, 1978; Lagakos et al., 2020). The key to their argument is that as income decreases, migrants may see their endeavour as unsuccessful, leading them to either return home or move to another place where they can achieve their goals.

However, income possesses inherent limitations as a utility measure when investigating whether floating migration is spurred by success or unmet expectations, we may ignore some facts in this case. Firstly, income might not inherently serve as the immigrant's primary yardstick to gauge the triumph of the migration venture. Individuals who migrate for family reunification or for the purpose of providing education to the subsequent generation illustrate instances where income inadequately encapsulates the myriad factors that contribute to the verdict of migration being triumphant or otherwise. Secondly, in contrast to life satisfaction, income can be stored and utilized subsequently in a different geographical setting. This particular characteristic of income gave rise to the perspective of the New Economics of Labor Migration (NELM) theory (Constant and Massey, 2002). Migrants who intend to work and amass savings in the host nation with the ultimate goal of returning to their home region will perceive the act of returning as a triumph. Nonetheless, it remains challenging to discern the extent to which either of the two mechanisms (moving due to perceived failure or due to success) plays a predominant role. Subjective well-being is characterized by its capability to assess attained expectations and emerges as a more suitable candidate to address this inquiry from a broader and more comprehensive perspective, as (Schiele, 2021) supports.

### 3.2.3 Life Satisfaction and Migration

From the preceding sub-sections, it is evident that subjective well-being stands as a dependable metric for assessing overall measurement, even when applied across diverse cultures and regions. As an indicator of realized aspirations, it assumes an even more pertinent role in the context of return or floating migration. Extensive research has delved into examining the correlation between SWB and migration dynamics.

From the perspectives of cross-country studies: As this result emerges in the initial phase of migration, the cause could be unrealistic expectations about the quality of life in the destination country. Fuchs-Schündeln and Schündeln (2009) find that post-migration life satisfaction improves for permanent migrants but remains unaltered for return migrants. Migrants who have permanent intentions and achieve positive outcomes by migrating to another country are more likely to experience a significant increase in life satisfaction, whereas temporary migrants are supposed to be less affected if their stated initial purpose was to stay only temporarily in the host country. Graham and Markowitz (2011) found that individuals with the intent to emigrate have a lower SWB. Cai et al. (2014) found that individuals with lower SWB have a higher propensity to emigrate. This holds true on the individual level and the country level. On the other hand, Bartram (2013) found the opposite pattern for

potential Eastern European migrants. Mara et al. (2013) show, in accordance with the theory of this paper, an increased propensity of return migration for low SWB Romanian migrants to Austria. Mrittika and Marina-Selini (2020) show that migrants who are happy with their life tend to stay permanently in Germany. Guedes Auditor and Erlinghagen (2021) find that compared with immobile German immigrants, first-time German immigrants have significantly higher SWB shortly after arrival in the host country.

Regarding internal migration, investigations into the correlation between migration and well-being using life satisfaction measurements are relatively scarce in the existing literature. Knight and Gunatilaka (2010), as well as Bartram (2011), employing cross-sectional data for both immigrants in the United States and internal migrants in China, generally observe that the life satisfaction of migrants tends to be lower than that of non-migrants. However, cross-sectional comparisons might be influenced by a selection bias stemming from disparities between migrants and non-migrants, irrespective of whether the analysis spans countries or internal migrations.

Addressing this concern, De Jong et al. (2002) utilize data from internal migrants within Thailand to challenge the assumptions of neoclassical immigration theory. Their findings indicate that emigration within Thailand is linked to a decline in post-immigration satisfaction, along with an increase in employment satisfaction but a decrease in satisfaction with the living environment following relocation. Nakazato et al. (2011) investigate German residential migrants, revealing that while housing satisfaction tends to rise post-migration, overall life satisfaction remains unchanged. Easterlin (2001); Easterlin and Angelescu (2009) propose that even if migration brings about objective improvements, it might not necessarily result in elevated life satisfaction if expectations also increase in tandem. Conversely, Melzer (2011) uncovers a different outcome for migrants moving from East to West Germany, identifying a positive long-term connection between migration and life satisfaction.

Exploring the British Household Panel Survey, Findlay and Nowok (2012); Nowok et al. (2013) ascertain that migrants' life satisfaction diminishes before the move and then rebounds at the time of migration. Notably, despite the greater social and psychological costs associated, long-distance migrants appear to be at least as content as their short-distance counterparts. Furthermore, Faßhauer and Rehdanz (2015) examines studies of East-West immigration following German reunification, which reveal a notable rise in life satisfaction solely among East-West permanent immigrants. Switek (2016) underscores that internal migration is linked to heightened short-to-medium-term life satisfaction for both working and non-working migrants. However, only working immigrants experience a substantial and sustained increase in life satisfaction, persisting six or more years after their migration.

In summary, while subjective well-being (SWB) presents attractive characteristics as a utility measure in the context of migration, the empirical findings concerning internal floating migration are somewhat ambiguous. Furthermore, immigrant attractiveness varies from person to person, depending on one's goals, values, and circumstances. There may be simultaneous effects between immigration decision-making and life satisfaction. This implies that there may be an interplay between immigration decisions and individuals' satisfaction with their lives, rather than a one-way causal relationship, as discussed above. Individuals' immigration decisions may affect their life satisfaction in their new environment. If a person achieves their career goals, improves their economic status, gains a better quality of life, etc. through immigration, they are likely to feel more satisfied. Conversely, life satisfaction may decrease. Individuals' life satisfaction may influence their decision to continue migrating or return to their place of origin, or to continue migrating. If a person has a good time in the new place, is well integrated into the society, has a satisfactory job, and enjoys a high quality of life, they may be more inclined to stay in the place of immigrant. Instead, return or continued immigration may be considered. Therefore, the relationship between immigration decisions and life satisfaction may be bidirectional. Migration decisions and life satisfaction may interact in a circular way. Therefore, in addition to strengthening the evidence on the relationship between internal immigration decision-making or immigration frequency and life satisfaction, this chapter will also demonstrate this issue from the perspective of simultaneous effects.

## 3.3   Data Description

The data for this study are derived from the China Family Panel Survey (CFPS), the China City Statistical Yearbook (CCS), and the China High-Resolution Emission Database (CHRED). CFPS is a nationally representative, annual longitudinal survey of Chinese communities, families, and individuals launched in 2010 by the Institute of Social Science Survey (ISSS) of Peking University, China. It is conducted every two years. The CFPS is designed to collect individual-, family-, and community-level longitudinal data in contemporary China. The CFPS focuses on the economic and non-economic well-being of the Chinese population, with a wealth of information covering such topics as economic activities, education outcomes, family dynamics and relationships, migration, and health. The CFPS sample covers 25 provinces/municipalities/autonomous regions, the target sample size is 16,000 households, and the survey objects include all family members in the sample

households[3]. CCS provides cities' characteristics where each individual allocates. CHRED provides the environmental factors that influence people's life satisfaction. The time period is chosen between 2012 and 2020.

The primary variables under investigation encompass life satisfaction and migration decisions. The assessment of life satisfaction transpires across all waves of the CFPS (2012–2020) via the query: "*How satisfied are you with your life in general?*" Responses are categorised on a 1 to 5 scale, with 1 signifying "very dissatisfied" and 5 indicating "very satisfied." We define the categorical variable "life satisfaction" as *LS*. It represents a person's life satisfaction during the observation period, regardless of their immigration status. Moreover, construct a first-order lagged term on *LS*, which is denoted as *Lag_LS*. Migration determinations are ascertained through individuals' self-reported data on their change in Hukou and residence within the CFPS. A person is categorised as a migrant if their municipality of residence changed during the analysed years, even including those who reported a different municipality, the value takes 1, otherwise, it is denoted as *Mig*.

CFPS is a very rich data set. Following the literature, see Section 3.2, from it, we can obtain individual-level demographic and social factor characteristics, including *Age* and its squared term (*Age_sq*), the job is in rural or urban (*Urban*), *Monthly_Income*, *Male*, *Religion*, *Education*, *Marry*, *Family_Tie*, and *Average Social Network Expense*. From the perspective of personal factors, I select variables such as male, religion, education, age, etc. to measure the individual characteristics of the labour force. Among them, *Male* is 1 if male, otherwise; *Religion* is 1 if Muslim, otherwise; *Age*, if male, the age range is 18-60, and female age range is 18-55, for a readable reason, the value of *Age_sq* is divided by 100; *Marry*, if married, is 1, otherwise. The reason is as follows, the decision to migrate may be influenced by the conditions experienced by the individual. The variable of education takes the value of an individual's highest year of schooling. Those who are older and less educated especially face high employment costs and limited employment opportunities in cities with large migrant populations, and therefore are more inclined to return to their hometowns (Zhao et al., 2023).

From the perspective of economic factors, in terms of monthly income, Blanchflower and Oswald (2004); Blanchflower (2009), using longitudinal and survey data from different countries, show a positive relationship between higher income levels and happiness. This effect becomes especially pronounced in transitional

---

[3]Regarding the issue of attrition sample non-traceability loss in CFPS: According to the CFPS data quality technical report each year, there are two major reasons for the loss of personnel and the inability to track them. One is random (new family personal genetic members or genetic family members; doubts, becoming a monk, joining the army, serving a sentence, etc. do not need to be tracked), and the other is the cause of death. Therefore, overall, the data does not have artificial selection or systematic instability, and the problem of selection bias can be ignored.

countries and situations involving interpersonal comparisons (Selezneva, 2011), therefore, income is more relevant to life satisfaction. The value of income takes logarithm form, which is symbolized as $ln\_Month\_Income$. On the other hand, historically, scholars have argued that employment is one of the most significant factors driving internal migration (Chen and Rosenthal, 2008; Yang et al., 2022), and some people will migrate if there are better jobs elsewhere (Chauvin et al., 2017; Zhao et al., 2023), the variable $Urban$ is closer to migration decision. The definition is if a person lives in municipalities directly under the Central Government (Beijing, Shanghai, Tianjin, Chongqing) and capital cities of each province and Shenzhen, the value takes 1, otherwise.

From the perspective of family and social factors, the closeness of the family relationship or how strongly it binds a person to one place will have an impact on a person's life satisfaction when he or she moves (Fan, 2011). To take this issue into account, to evaluate how strongly an individual ties with his or her family, I use how many evenings per week you eat with your family ("family" means members of the household who live with you and does not include relatives who do not live with you). The value takes 0–7, the symbol is $Family\_Tie$. When a person moves, the strength of social links or the amount of investment a person makes in a city will have an impact on his or her life satisfaction. To capture this, the personal average social network expense could be an appropriate proxy to test, the value takes log form and symbolises as $ln\_AVE\_Social\_Network$.

Moreover, I control the following city-dimensional variables: in society, demographic, macroeconomic, and environmental factors influence life satisfaction (Rajani et al., 2019); therefore, it is relevant to both in different contexts. Research has found that migration is a density-dependent function. As the population density approaches the carrying capacity, the amount of migration will decrease (Ossandon and Castro-Santis, 2018); Population Density is defined as the number of people in a city divided by the area of the city. Secondly, for the variable of the Unemployment Rate, Ransom (2016) shows that there is a relationship between immigration decisions and the unemployment rate; however, there is no clear literature showing that the unemployment rate at the city level will lead to changes in life satisfaction. Industrial Structure is defined as the ratio of manufacturing GDP to service industry GDP, where changes in local industries induce interregional migration and intraregional migration, with different impacts on rural-urban, urban-rural, urban-urban, and rural-rural migration (Matsukawa, 1991). Furthermore, immigration and health are interrelated. Immigration can cause public health problems (Jeffrey, 1987). The relationship between immigration and health is complex, with both positive and negative effects. Including immigrants' health,

public health, and healthcare systems in this issue (Vearey et al., 2020). So I control the number of hospitals in a city, *Number of Hospitals*. At the same time, a person's immigration goal may be based on the education of the next generation (Mosselson and Chinkondenji, 2020), so I choose to control the number of primary and secondary schools in a city as a proxy variable for the investigation, Number of Schools. Because of the scale issue, the value of the Number of Hospitals and the Number of Schools is divided by 1000.

Furthermore, I also control for the influence of environmental factors on life satisfaction, Silva et al. (2012) showed that environmental quality has a significant effect on life satisfaction, the magnitude of which is about half of the self-reported health status, the environment is an external factor affecting life satisfaction (Andreenkova and Andreenkova, 2019); moreover, nature experiences, including visiting different natural spaces and viewing natural landscapes from windows, were positively associated with higher life satisfaction and well-being (Chang et al., 2020). Therefore, I control the rest of the relevant environmental factors, Carbon Emissions, Built City Area Green Coverage Rate, Industrial Sulphur Dioxide Emission, Air Quality Index (AQI), PM2.5, and PM10. The value of Carbon Emissions, and Industrial Sulphur Dioxide Emission takes the logarithm form.

The data description is in Table 3.1. From Table 3.1, we could see on average, about 21.9% of observations made moving decisions in each observed period; in China, people's average life satisfaction is 3.7. Figure 3.1 shows the distribution of life satisfaction.

Figure 3.1: Distribution of Life Satisfaction
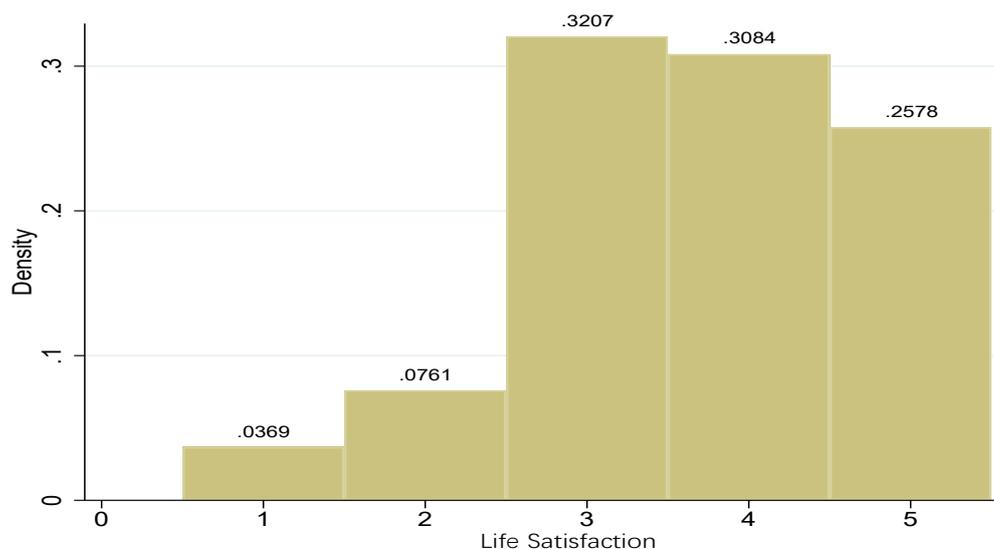
Table 3.1: Data Description

|  | Obs | Mean | SD | Min | Max |
|---|---|---|---|---|---|
| Mig | 48,344 | 0.207 | 0.405 | 0 | 1 |
| LS | 46,054 | 3.725 | 1.045 | 1 | 5 |
| Lag_LS | 44,752 | 3.497 | 1.102 | 1 | 5 |
| Age | 48,344 | 45.489 | 12.081 | 16 | 65 |
| Age_sq | 48,344 | 22.152 | 10.630 | 2.560 | 42.250 |
| Edu | 47,952 | 7.363 | 4.747 | 0 | 22 |
| Marry | 47,603 | 0.884 | 0.321 | 0 | 1 |
| Male | 48,344 | 0.490 | 0.500 | 0 | 1 |
| Family_Tie | 44,170 | 6.379 | 1.525 | 1 | 7 |
| Popdensity | 42,058 | 458.573 | 367.761 | 0 | 2648.110 |
| Unemployrate | 40,700 | 0.139 | 0.339 | 0.002 | 2.955 |
| Industrialratio | 42,058 | 1.032 | 0.554 | 0.006 | 5.297 |
| Religion | 48,344 | 0.018 | 0.134 | 0 | 1 |
| Hospital | 42,058 | 0.209 | 0.211 | 0.013 | 2.168 |
| School | 42,058 | 1.127 | 0.685 | 0 | 4.312 |
| Urban | 48,344 | 0.478 | 0.500 | 0 | 1 |
| Month_Income | 43,411 | 3959.389 | 2616.535 | 0 | 150000 |
| AVE_Social_Network | 48,080 | 4.463 | 16.641 | 0 | 833.333 |
| CO2 | 46,536 | 40.669 | 42.766 | 1.754 | 226.676 |
| SO2 | 42,058 | 127527.500 | 177542.390 | 517 | 991959 |
| AQI | 41,104 | 97.875 | 51.957 | 18 | 352.250 |
| PM2_5 | 39,904 | 61.249 | 46.574 | 11 | 297.925 |
| PM10 | 41,104 | 105.850 | 67.252 | 18 | 461.358 |
| Greenrate | 42,058 | 38.607 | 9.887 | 0.390 | 158.110 |

Note: The value of $Age\_sq$ is divided by 100, and the values of $Hospital$, $School$ are divided by 1000.

## 3.4   Methodology

### 3.4.1   Baseline Modeling

The dependent variable in Equation (3.1) is $Mig$, which has values between 0 and 1, so I use the panel Probit model to look into the relationship between lagged self-reported life satisfaction and labour mobility. The panel Probit model ensures that predicted probabilities stay within this range and capture the non-linear nature of the relationship. Additionally, it allows for controlling unobserved individual-specific effects, improving the accuracy of the estimates by accounting for both temporal changes and individual heterogeneity. In Equation (3.2), given that life satisfaction is categorical, the appropriate model for assessing the impact of these factors on life satisfaction is the panel-ordered Probit model. Because the categories have a natural order but the intervals between them are not necessarily equal. This model estimates the probability that an observation falls into each category, accounting for the ordered nature of the life satisfaction variable.

Equation (3.1) encapsulates the function of labours' mobility decisions ($Mig$), which is influenced by the lag of life satisfaction ($Lag\_LS$). Additionally, we account for control variables at the demographic ($\boldsymbol{X_{1i,t}}$) and city ($\boldsymbol{City_{1c,t}}$) levels. In particular, $\boldsymbol{X_{1i,t}}$ includes the individual's age and its quadratic term, gender, year of schooling, family tie, whether the job is in the city, religious and marital status; $\boldsymbol{City_{1c,t}}$ includes the logarithm of population density, the natural value of a city's unemployment rate, the industrial ratio, and the number of hospitals and schools; $\boldsymbol{\alpha_2}$ and $\boldsymbol{\alpha_3}$ are the vectors of corresponding coefficients. $\mu_i$ and $\gamma_t$ represent individual and time effects, respectively. $\Phi$ is the cumulative distribution function of standard normal distribution.

$$
\begin{aligned}
Pr\left(Mig_{i,t}|Lag\_LS_{i,t}\right) =& \Phi(\alpha_0 + \alpha_1 Lag\_LS_{i,t} + \boldsymbol{\alpha_2}\boldsymbol{X_{1i,t}} \\
&+ \boldsymbol{\alpha_3}\boldsymbol{City_{1c,t}} + \mu_i + \gamma_t)
\end{aligned}
\tag{3.1}
$$

$$
LS_{i,t} = \begin{cases}
1 & \text{if } LS_{i,t} \leq LS_1 \\
2 & \text{if } LS_1 < LS_{i,t} \leq LS_2 \\
3 & \text{if } LS_2 < LS_{i,t} \leq LS_3 \\
4 & \text{if } LS_3 < LS_{i,t} \leq LS_4 \\
5 & \text{if } LS_4 < LS_{i,t}
\end{cases}
, where \ LS_{i,t} = 1,2,3,4,5
$$

$$Pr\left(LS_{i,t} \leq j\right) = \Phi(\theta_j + \eta_i + \lambda_t - \beta_0 - \beta_1 Mig_{i,t} - \boldsymbol{\beta_2 X_{2i,t}} \\ - \boldsymbol{\beta_3 City_{2c,t}}) \tag{3.2}$$

Equation (3.2) represents the relationship between life satisfaction and labour mobility decisions. In this equation, I incorporate the labours' mobility decision ($Mig$) to test our research hypothesis. Furthermore, we consider control variables at both the demographic ($\boldsymbol{X_{2i,t}}$) and city ($\boldsymbol{City_{2c,t}}$) levels. In particular, $\boldsymbol{X_{2i,t}}$ includes the individual's age and its quadratic term, gender, year of schooling, the logarithm of monthly income, average spending monthly network; $\boldsymbol{City_{2c,t}}$ includes logarithm of Carbon Emissions, Built City Area Green Coverage Rate, Industrial Sulphur Dioxide Emission, AQI, PM2.5, and PM10. $\boldsymbol{\beta_2}$ and $\boldsymbol{\beta_3}$ are the vectors of corresponding coefficients. $\eta_i$ and $\lambda_t$ represent individual and time effects, respectively. $\theta_j$ is a threshold parameter, which represents the boundary value at which the probability of $LS_{i,t}$ taking value j is zero. $\Phi$ is the cumulative distribution function of standard normal distribution.

### 3.4.2 Addressing Endogeneity

As I discuss in the Section 3.A.2.1, endogeneity issues exist between life satisfaction and mobility decision has reverse causality. In Equation (3.1), I employ lagged life satisfaction. Intuitively, this should mitigate concerns about endogeneity between the migration decision in the base period and life satisfaction. However, it is important to account for the fact that life satisfaction is a highly persistent variable. This persistence means that life satisfaction might be stable over time, reflecting past experiences, which could still introduce complexities in analysing its relationship with migration decisions. Therefore, I have this consideration that, in a short period, it builds up over time based on the history of the last period; therefore, an individual would not, or it would be hard to, change their subjective well-being based on an isolated event. Moreover, the decision to migrate or not may also be influenced by a decrease in life satisfaction within their current city at a given time. This interdependence establishes a cyclical relationship between these two factors, indicating a causal connection between them. Therefore, since I notice the endogeneity problem in regressions, I use the instrumental variable method to solve it. To conduct the results, I employ the conditional mixed process (CMP).

Due to the absence of TSLS panel Probit and panel Ordered-Probit estimation command in STATA, an alternative strategy is adopted by using the CMP which is suggested by (Roodman, 2009, 2011, 2023). Given that one of the dependent

variables is binary (mobility decision) and the other (life satisfaction) is ranked and ordered, the Conditional Mixed Process (CMP) undertakes the estimation of a system comprising seemingly unrelated equations. This methodology permits the endogenous variables (lagged life satisfaction and mobility decision) to feature as explanatory variables on the right-hand side of the opposing equation. The key advantage of CMP over independent equation estimation or two-stage least squares (2SLS) lies in the fact that the lagged life satisfaction concerning migration enters the migration intention equation solely as an explanatory and categorical variable, without accounting for unobservable factors affecting life satisfaction. Conversely, through CMP, the predicted value of life satisfaction is taken into account, aligning with the other equation. Typically, this approach yields unbiased and more efficient estimates, particularly under the assumption of normally distributed error terms (Mara et al., 2013).

To Equation (3.1), a more conservative approach is adopted, employing an individual's average one-way commute time and weekly overtime work hours as an instrumental variable. Firstly, the impact of overtime work on life satisfaction is multifaceted. Overtime work is generally associated with increased work stress, fatigue, and work-family interference, involuntary overtime significantly decreases happiness, life satisfaction, and health status, and increases depressive symptoms(Golden and Wiens-Tuers, 2006; Ahmad, 2016; Yang et al., 2023). Furthermore, the rationale behind choosing commute time as the second instrumental variable lies in empirical research demonstrating that longer commutes between residence and work correspond to increased stress levels and lower well-being reported (Ettema et al., 2010; Wener and Evans, 2011). My arguments for using the instruments are as follows. To overtime work, according to the China Labour Trends Report, overtime work is a common trend among the Chinese workforce, with the average hours reaching 7.1. This indicates in China overtime working is relatively common. To commute time, while previous literature primarily focuses on countries like Europe and the United States where commuting may occur across cities, leading individuals to relocate closer to their workplaces to minimise commuting time and costs, the current context in China typically involves individuals working and living in the same city. Furthermore, T-tests conducted on samples exhibiting movement in every observed period reveal no statistically significant difference between the two years of moving of commute time or overtime work [4]. These suggest that commute time and work overtime do not significantly impact individuals' mobility decisions in this context. Therefore, Equation (3.1)'s 2SLS is re-written as in follow Equation (3.3) and Equation (3.4):

---

[4]The T-tests results tables are in Tables 3.13 to 3.20

Equation (3.3) is the First Stage of Instrumental Variable Process to Equation (3.4),

$$
\begin{aligned}
Pr\left(\widehat{Lag\_LS}_{i,t} \leq j\right) = \Phi(\vartheta_j + \upsilon_i + \iota_t - \varrho_0 - \varrho_1 Commute\_time_{i,t} \\
- \varrho_2 Overtime\_time_{i,t} - \boldsymbol{\varrho_3 X_{1i,t}} - \boldsymbol{\varrho_4 City_{1c,t}})
\end{aligned}
\tag{3.3}
$$

$$
\begin{aligned}
Pr\left(Mig_{i,t}|\widehat{Lag\_LS}_{i,t}\right) = \Phi(\nu_0 + \nu_1 \widehat{Lag\_LS}_{i,t} + \boldsymbol{\nu_2 X_{1i,t}} \\
+ \boldsymbol{\nu_3 City_{1c,t}} + \rho_i + \varsigma_t)
\end{aligned}
\tag{3.4}
$$

To Equation (3.2), the instrumental variable is derived from the work of Combes et al. (2015), the spherical distances between the host city and seaport cities. The rationale behind this instrument is rooted in the notion that historically significant cities, established as immigrant centers centuries ago, continue to attract modern immigrants. However, attractiveness is not necessarily tied to changes in life satisfaction in the current period. As noted by Wu and Gaubatz (2013), China's cities underwent transformative changes during major historical events, including the establishment of foreign treaty ports as part of the "unequal treaties" signed with Western countries in the late Qing Dynasty. Identifying key cities from the late 19th and early 20th centuries that evolved into major economic centers after the opening of 48 treaty ports between 1842 and 1920 serves as the basis for this instrument. Jia (2014) employs whether the city was a treaty port post-Opium War as an instrument variable. Moreover, 14 coastal and port cities have newly opened since China's Reform and Opening-up. The process first involves using dummy variables to identify these cities. Economic geography principles state that "centrality" in an economy influences productivity, wages, and employment, attracting migrants. Proximity to historical or coastal port cities can act as a substitute for being physically close, as interactions remain feasible within a reasonably short distance (Combes et al., 2015). Since the treaty ports are, to a large extent, the same as the seaports, they also add the average distance to the seaports as an instrument. However, I try to improve this average distance in my research, as this variable may underestimate the geographic advantage of those cities near the seaports. To overcome this shortcoming, I use the minimum distance instead. Second, I calculate the minimum distance of the spherical distances between the target city and seaport cities using (Vincenty, 1975) method. Therefore, Equation (3.2)'s 2SLS is re-written as in follow Equation (3.5) and Equation (3.6):

Equation (3.5) is the First Stage of Instrumental Variable Process to Equation (3.6),

$$Pr\left(\widehat{Mig_{i,t}}|Distance_{i,t}\right) = \Phi(\kappa_0 + \kappa_1 \text{Distance}_{i,t} + \boldsymbol{\kappa_2 X_{2i,t}} + \tag{3.5}$$
$$\boldsymbol{\kappa_3 City_{2c,t}} + \zeta_i + \xi_t)$$

$$Pr\left(LS_{i,t} \leq j\right) = \Phi(\tau_j + \delta_i + \varepsilon_t - \pi_0 - \pi_1 \widehat{Mig_{i,t}} - \boldsymbol{\pi_2 X_{2i,t}} \tag{3.6}$$
$$- \boldsymbol{\pi_3 City_{2c,t}})$$

To address the absence of instrumental variable specification tests specifically tailored for nonlinear estimation, I adopt the methodology proposed by Di Maria et al. (2018) to address this concern. Their study utilizes the conditional mixed process to conduct instrumental variable two-stage regression analyses, with dependent variables ranging in binary and categorical. Their findings report results from under-identification, weak identification, and over-identification tests for the 2SLS method. However, their practical approach involves assuming linearity in the dependent variable and reporting corresponding specifications through linear 2SLS regression. Their rationale is that if the model meets the requirements under the linear specification, it is likely to meet the requirements under the nonlinear specification as well. I intend to include these sections of the specifications in Tables 3.45 and 3.47 of my study for reference and transparency in methodology.

## 3.5 Results

The baseline results are served in Tables 3.2 and 3.5, and columns 1 to 4 show robust results when gradually adding individual demographic characteristics control variables, corresponding to Equations (3.1) and (3.2) respectively. Either FE estimation is susceptible to bias due to endogeneity concerns. To address this, the CMP-IV approach is employed. The results in the fifth and sixth columns correspond to Equations (3.3) and (3.5) and Equations (3.4) and (3.6), respectively. Column 5 is the second-stage regression results under the CMP setting. Column 6, shows the first stage of CMP-IV results. Moreover, due to each estimation's function form, the coefficients do not have a direct economic interpretation; therefore, I compute the results from the coefficient level to the average marginal effects and conditional marginal effects. The marginal effect results are shown in Table 3.4, and the conditional marginal effect results are shown in Table 3.6.

We can see from Table 3.2, that as I gradually add personal characteristics control variables, the coefficient levels for our variable of interest, lagged life satisfaction, tend to be robust. A negative correlation exists between lagged life

satisfaction levels and migration decisions at the 1% statistical level. At the overall level, when the lagged life satisfaction level increases, the probability of the labour force making a migration decision decreases by 6.3 p.p., in Table 3.4 column 1. This is also consistent with the conclusions drawn in the literature. Some studies suggest higher life satisfaction reduces the likelihood of migration decisions, while other studies indicate it may increase the intention to migrate. Individuals reporting high life satisfaction are generally less likely to intend to migrate or relocate, even when controlling for education and other background factors, suggesting that satisfied individuals prefer stability (Luhmann et al., 2013; Ivlevs, 2014; Ivlevs et al., 2014; Chindarkar, 2014; Ossandon and Castro-Santis, 2018). In Table 3.1, I report that the mean of lagged life satisfaction is approximately 3.5. In the CFPS, life satisfaction is defined as a range of 1 to 5. We can say that in the overall sample, the life satisfaction level of the Chinese labour force is relatively high, which is consistent with the estimation results.

As we discuss, life satisfaction is an ordinal variable and everyone reports their level of subjective well-being. Therefore, in addition to focusing on whether the overall lagged life satisfaction is the driving factor for people to migrate, we also need to consider the group situation under each life satisfaction level. So in Table 3.3, I redo the CMP-IV estimates, treating lagged life satisfaction as a grouping variable to get separate regression results for each level. And the marginal effects are reported in Table 3.4 column 2. We can see at the coefficient level when the individual's LS is 2, lagged life satisfaction and mobility decision have a positive association at 1% statistical significance. It means at the marginal level individuals' lagged life satisfaction is 2, and they are more likely to make mobility decisions by 29.0 p.p., compared with the baseline when life satisfaction is 1. Lower-lagged life satisfaction plays a significant role in influencing migration decisions, as evidenced by various research studies(Graham and Markowitz, 2011; Cai et al., 2014). Similarly, studies using Eurobarometer Survey data from Central Eastern and Western European countries indicate that dissatisfied individuals have a higher intention to migrate, with life satisfaction acting as a mediator between macroeconomic conditions and migration decisions (Otrachshenko and Popova, 2012; Erlinghagen, 2016).

The observation that individuals with a life satisfaction score of 2 are more likely to move than those with a score of 1 seems to contradict the general notion that less satisfied individuals are more inclined to relocate. One possible explanation for this anomaly can be found in Sen Sen (1990), which highlights the importance of real freedoms or opportunities individuals have to pursue lives they value. According to Sen, a person's capabilities depend not only on their aspirations but also on their actual ability to act on those aspirations, which can be constrained by various personal

Table 3.2: Estimates Equation 1 using single equation

| VARIABLES | (1) xtprobit Mig | (2) xtprobit Mig | (3) xtprobit Mig | (4) xtprobit Mig | (5) CMP-IV Mig | (6) CMP-IV Lag_LS |
|---|---|---|---|---|---|---|
| Overtime_work | | | -0.002** | | | |
| | | | (0.001) | | | |
| Commute_time | | | -0.003*** | | | |
| | | | (0.001) | | | |
| Lag_LS | -0.277*** | -0.322*** | -0.317*** | -0.375*** | -0.250*** | |
| | (0.028) | (0.047) | (0.030) | (0.085) | (0.010) | |
| Age | | -0.806** | -0.720*** | -0.875*** | -0.156*** | -0.017*** |
| | | (0.379) | (0.059) | (0.075) | (0.005) | (0.006) |
| Age_sq | | 0.609*** | 0.575*** | 0.658*** | 0.139*** | 0.028*** |
| | | (0.179) | (0.077) | (0.091) | (0.006) | (0.007) |
| Edu | | | 0.015 | 0.045 | -0.009*** | 0.009*** |
| | | | (0.049) | (0.038) | (0.002) | (0.002) |
| Marry | | | | 0.197 | 0.021 | 0.201*** |
| | | | | (0.356) | (0.030) | (0.032) |
| Male | -0.323 | 1.466 | 1.079** | 1.256*** | 0.280*** | -0.044** |
| | (0.427) | (1.162) | (0.535) | (0.309) | (0.017) | (0.018) |
| Family_Tie | -0.056** | -0.048* | -0.052** | -0.051 | -0.026*** | 0.009* |
| | (0.025) | (0.029) | (0.025) | (0.055) | (0.005) | (0.005) |
| Urban | -0.369*** | -0.352*** | -0.341*** | -0.279 | -0.190*** | 0.114*** |
| | (0.080) | (0.118) | (0.089) | (0.230) | (0.018) | (0.021) |
| $N$ | 37,094 | 37,094 | 36,818 | 36,680 | 36,680 | 36,680 |
| City Controls | YES | YES | YES | YES | YES | YES |
| Fixed Effects | YES | YES | YES | YES | YES | YES |

Note: Robust standard errors in parentheses. ***Significant at 1%; **Significant at 5%; *Significant at 10%. Full tables are in Tables 3.21 and 3.22. Fixed effects include individual and time effects. We report the OLS results in Table 3.33. The intuition is that preliminary results based on the span of data type methods can still show us the signs associated with our coefficients. I report the IV specifications in Tables 3.45 and 3.46, by combining with Tables 3.2 and 3.22, the coefficient is statistically significant at 1% level, which means that the instrumental variables are significant in predicting the endogenous explanatory variables in the first stage. Moreover, for the Sanderson-Windmeijer multivariate F test of excluded instruments is 15.35 ($p-value = 0.000$) the IV F-stat is 110.51, and Anderson LM statistic is 30.676 ($p-value = 0.000$), which indicate the IVs pass the underidentification test. Cragg-Donald Wald F statistic is equal to 15.348, and according to Hausman et al. (2005); Stock and Yogo (2005), the 10% maximal size is 8.68, the 15.348 is greater than Stock-Yogo weak ID test critical value, therefore, the joint instrumental variable passes weak identification test. Lastly, Hansen's test fails to reject the null that the instruments are valid ($J-test = 0.069$, $p-value = 0.793$), which means our joint instrumental variable is not over-identified.

Table 3.3: Estimates Equation 1 using single equation by treating the lagged LS as ordered variable

| VARIABLES | (1) CMP-IV Mig | (2) CMP-IV Lag_LS |
|---|---|---|
| Overtime_work | -0.002** | |
| | (0.001) | |
| Commute_time | -0.003*** | |
| | (0.001) | |
| Lag_LS=2 | 0.928*** | |
| | (0.050) | |
| Lag_LS=3 | -0.060 | |
| | (0.050) | |
| Lag_LS=4 | -0.158*** | |
| | (0.053) | |
| Lag_LS=5 | -0.162*** | |
| | (0.057) | |
| Age | -0.149*** | -0.017*** |
| | (0.005) | (0.006) |
| Age_sq | 0.132*** | 0.028*** |
| | (0.006) | (0.007) |
| Edu | -0.007*** | 0.009*** |
| | (0.002) | (0.002) |
| Marry | 0.018 | 0.201*** |
| | (0.031) | (0.032) |
| Male | 0.266*** | -0.044** |
| | (0.017) | (0.018) |
| Family_Tie | -0.027*** | 0.009* |
| | (0.005) | (0.005) |
| Urban | -0.181*** | 0.114*** |
| | (0.019) | (0.021) |
| N | 36,680 | 36,680 |
| City Controls | YES | YES |
| Fixed Effects | YES | YES |

Note: Robust standard errors are in parentheses. ***Significant at 1%; **Significant at 5%; *Significant at 10%. The full tables are in Tables 3.23 and 3.24. Fixed effects include individual and time effects. The IV specifications see in Table 3.2.

and socio-economic factors. Individuals with the lowest life satisfaction (LS=1) may face significant barriers, such as financial difficulties, limited access to information, or a lack of social support, restricting their mobility despite a strong desire to improve their situation. In contrast, those with slightly higher life satisfaction (LS=2) may have marginally better resources or support systems, enabling them to act on their desire to move. Thus, the higher mobility observed among individuals with LS=2 compared to LS=1 likely reflects differences in capabilities rather than a difference in the desire to relocate. This interpretation aligns with Sen's perspective, which emphasizes that true freedom requires both the aspiration to achieve certain outcomes and the genuine opportunity to realise them.

Furthermore, longitudinal data analysis has revealed a notable decrease in life satisfaction in the years leading up to migration events, indicating a correlation between declining life satisfaction and the decision to emigrate (Erlinghagen, 2016). Furthermore, longitudinal data from the German Socio-Economic Panel reveals a significant decrease in life satisfaction leading up to emigration events, suggesting a link between lower life satisfaction and migration choices (Otrachshenko and Popova, 2012). Nevertheless, if an individual's previous life satisfaction score is 3 or higher, compared with baseline, they exhibit negative connections. However, when the score is exactly 3, it lacks statistical significance. The marginal results show that if the lagged life satisfaction is equal to 4 or 5, individuals are less likely to make moving decisions by 3.7 p.p. and 3.8 p.p. than life satisfaction is very low, respectively. The findings are aligned with the literature, research indicates that individuals with high life satisfaction are less likely to migrate, as satisfaction with their current circumstances reduces the desire for change (Laurinavičius et al., 2021; Koroļeva, 2022).

Table 3.4: The Marginal Effects of the Variables of Interest on Mobility Decisions

| | (1) | | (2) | |
| --- | --- | --- | --- | --- |
| | Mig | | Mig | |
| | b | se | b | se |
| Lag_LS | -0.063*** | 0.003 | | |
| Lag_LS=2 | | | 0.290*** | 0.013 |
| Lag_LS=3 | | | -0.014 | 0.012 |
| Lag_LS=4 | | | -0.037*** | 0.013 |
| Lag_LS=5 | | | -0.038*** | 0.014 |
| Age | -0.039*** | 0.001 | -0.036*** | 0.001 |
| Age_sq | 0.035*** | 0.001 | 0.032*** | 0.001 |
| Edu | -0.002*** | 0.001 | -0.002*** | 0.001 |
| Marry | 0.005 | 0.008 | 0.004 | 0.007 |
| Male | 0.071*** | 0.004 | 0.064*** | 0.004 |
| Family_Tie | -0.006*** | 0.001 | -0.006*** | 0.001 |
| ln_Pop_density | -0.018*** | 0.003 | -0.015*** | 0.003 |
| Unemployrate | -0.018*** | 0.007 | -0.016** | 0.006 |
| Industrialratio | 0.012** | 0.005 | 0.013*** | 0.005 |
| Religion | 0.004 | 0.020 | -0.000 | 0.019 |
| Hospital | -0.042*** | 0.011 | -0.039*** | 0.011 |
| School | 0.024*** | 0.003 | 0.020*** | 0.003 |
| Urban | -0.048*** | 0.005 | -0.043*** | 0.005 |
| N | 36,680 | 36,680 | | |

Note: Robust standard errors in parentheses. ***Significant at 1%; **Significant at 5%; *Significant at 10%. In column (2), the marginal effects of Lag_LS are reported by treating the lagged LS as the ordered variable.

Table 3.5: Estimates Equation 2 using single equation

| VARIABLES | (1) xtoprobit LS | (2) xtoprobit LS | (3) xtoprobit LS | (4) xtoprobit LS | (5) CMP-IV LS | (6) CMP-IV Mig |
|---|---|---|---|---|---|---|
| Distance | | | -0.002*** | | | |
| | | | (0.000) | | | |
| Mig | 0.057** | 0.037* | 0.033* | 0.027* | 0.308*** | |
| | (0.023) | (0.025) | (0.025) | (0.025) | (0.048) | |
| Age | -0.044*** | -0.045*** | -0.065*** | -0.032*** | -0.168*** | |
| | (0.005) | (0.005) | (0.006) | (0.005) | (0.006) | |
| Age_sq | 0.057*** | 0.056*** | 0.078*** | 0.042*** | 0.147*** | |
| | (0.006) | (0.006) | (0.006) | (0.005) | (0.007) | |
| Edu | | -0.011*** | -0.011*** | -0.009*** | -0.005** | |
| | | (0.002) | (0.002) | (0.002) | (0.002) | |
| Marry | | 0.311*** | 0.242*** | -0.050 | | |
| | | (0.031) | (0.022) | (0.031) | | |
| Male | -0.044** | -0.053*** | -0.038** | -0.032* | -0.032** | 0.324*** |
| | (0.018) | (0.018) | (0.019) | (0.019) | (0.013) | (0.019) |
| Lag_LS | 0.150*** | 0.146*** | 0.151*** | 0.149*** | 0.336*** | -0.114*** |
| | (0.009) | (0.009) | (0.010) | (0.012) | (0.006) | (0.009) |
| ln_Month _Income | 0.007 | 0.019*** | 0.021*** | 0.019*** | 0.017*** | -0.003 |
| | (0.005) | (0.006) | (0.006) | (0.006) | (0.005) | (0.008) |
| ln_AVE_ Social_Network | -0.003 | 0.004 | 0.012* | 0.011* | 0.013** | -0.026*** |
| | (0.007) | (0.007) | (0.007) | (0.007) | (0.006) | (0.009) |
| N | 30,847 | 30,847 | 30,579 | 30,579 | 30,579 | 30,579 |
| City Controls | YES | YES | YES | YES | YES | YES |
| Fixed Effects | YES | YES | YES | YES | YES | YES |

Note: Robust standard errors in parentheses. ***Significant at 1%; **Significant at 5%; *Significant at 10%. The full tables are in Tables 3.25 and 3.26. Fixed effects include individual and time effects. The constant term is from the first stage of IV estimation. We report the OLS results in Table 3.34. The intuition is that preliminary results based on the span of data type methods can still show us the signs associated with our coefficients. I report the IV specifications in Tables 3.47 and 3.48, by combining with Table 3.5, the coefficient is statistically significant at 1% level, which means that the instrumental variables are significant in predicting the endogenous explanatory variables in the first stage. Moreover, for the Sanderson-Windmeijer multivariate F test of excluded instruments is 83.250 ($p-value = 0.000$) the IV F-stat is 226.88, and Anderson LM statistic is 83.079 ($p-value = 0.000$), which indicate the IVs pass the underidentification test. Cragg-Donald Wald F statistic is equal to 83.252, and according to Stock and Yogo (2005), the 10% maximal size is 8.68, the 15.348 is greater than the critical value, therefore, the joint instrumental variable passes the weak identification test.

Table 3.6: The Marginal Effects of the Variables of Interest on Life Satisfaction

| | (1) | |
| | LS | |
| | b | se |
|---|---|---|
| i.Mig=1 | | |
| LS=1 | -0.021*** | 0.003 |
| LS=2 | -0.027*** | 0.004 |
| LS=3 | -0.061*** | 0.009 |
| LS=4 | 0.013*** | 0.002 |
| LS=5 | 0.095*** | 0.015 |
| N | 30,579 | |

Note: Robust standard errors in parentheses. ***Significant at 1%; **Significant at 5%; *Significant at 10%. Here reports when $Mig = 1$, the marginal effects respect each level of life satisfaction. The full tables of the marginal effect of each level of life satisfaction are in Tables 3.35 to 3.39.

On the other hand, I will answer my second research question: does mobility increase life satisfaction? From Table 3.5, we observe that as the personal characteristics control variables are gradually included and under the setting of fixed effects of panel ordered Probit, the coefficient level steadily stabilises. At the overall level, life satisfaction is favourably associated with the choice to relocate. In Table 3.6, I present the conditional marginal effect for each outcome when Mig = 1. The coefficients shown are the conditional marginal effects of migration (Mig = 1) on the probability of being in each level of life satisfaction (LS = 1, 2, 3, 4, 5)[5][6][7]. Migration leads to a 2.1 p.p. decrease in the probability of being at the lowest level of life satisfaction. Similarly, it results in a 2.7 p.p. decrease in the probability of being at the second lowest level of life satisfaction, significant at the 1% level, and a 6.1 p.p. decrease in the probability of being at the third level of life satisfaction, also significant at the 1% level. Conversely, migration leads to a 1.3 p.p. increase in the probability of being at the second highest level of life satisfaction, significant at the 1% level, and a 9.5 p.p. increase in the probability of being at the highest level of life satisfaction, significant at the 1% level. It is important to note that the sum of these marginal effects is zero. This evidence suggests that people who migrate are less likely to allocate lower life satisfaction and more likely to allocate higher life satisfaction than those who do not migrate, indicating that migration decision has a positive association with life satisfaction. As demonstrated in the literature,

---

[5]Because the dependent variable in panel-ordered Probit is assumed a latent continuous variable that is segmented into numerous ordered categories using a series of thresholds. The impact of the explanatory variable on the underlying continuous variable is not linear, hence the conditional average effect provides a more precise representation of this non-linear relationship.

[6]Life satisfaction is an ordinal variable with values ranging from 1 to 5, where 1 indicates the lowest level of life satisfaction and 5 indicates the highest level.

[7]Mobility decision is a binary variable indicating whether an individual has migrated (Mig = 1) or not (Mig = 0).

combined with my data structure of every two years panel data, internal migration often leads to a short to medium-term increase in life satisfaction (De Jong et al., 2002; Melzer, 2011; Findlay and Nowok, 2012; Nowok et al., 2013; Bartram, 2015; Switek, 2016).

## 3.6    Robustness Checks

In additional analysis, I will answer two research sub-questions. The first is whether migrants' life satisfaction improves when they choose to leave big cities; the second is to explore the heterogeneous effects of education level on migration decisions through life satisfaction by constructing interaction terms.

I first re-divide the total sample, identifying which migrant individuals in the matching sample decided to leave capital cities (including Beijing, Shanghai, Tianjin, Chongqing, provincial capitals, etc.).   I rename the variable name from $Mig$ to $Return\_Mig$[8]. Using the subsample, I estimate whether deciding to leave big cities improved their life satisfaction, the estimation results are in Tables 3.7 and 3.9. The marginal effects of lagged life satisfaction on return migration decision are shown in Table 3.8 The conditional marginal effect results are shown in Table 3.10.

Again, I start by investigating how life satisfaction influences migration decisions among return migrant groups.  From Table 3.7, in columns (2) and (4), past life satisfaction is positively associated with the migration decision of return migrants, both at the overall mean level of lagged life satisfaction and at each level of it. At the marginal effect level, on average compared to the baseline lagged LS=1, an increase in lagged life satisfaction is associated with a 7.4 p.p.  higher probability of deciding to return to migrate.  This suggests that individuals who were more satisfied with their life in the past are significantly more likely to choose to return. When we break down the lagged life satisfaction to each level, we can clearly see that when Lag_LS = 2, there are 10.6 p.p.  more likely to return migrate compared to the baseline group; to Lag_LS = 3, there are 20.0 p.p.  more likely to return migrate; to Lag_LS = 4, there are 33.7 p.p.  more likely to return migrate; lastly, to Lag_LS = 5, there are 53.8 p.p. more likely to return migrate. It shows a clear and increasing positive impact on the likelihood of return migration as the level of past life satisfaction increases, also further suggesting robust evidence that higher past life satisfaction strongly influences the decision to return.

The above results may be due to the interaction of complex factors other than

---

[8]Clarification in essence, the definition of the return migration decision is the same as the migration decision definition in the previous article, but for ease of understanding, I rename the variable name to $Return\_Mig$.

Table 3.7: the Estimation of Lagged Life Satisfaction on Returning Migrations

| VARIABLES | (1) xtprobit Return_Mig | (2) CMP-IV Return_Mig | (3) CMP-IV Lag_LS | (4) CMP-IV Return_Mig | (5) CMP-IV Lag_LS |
|---|---|---|---|---|---|
| Commute_time | | | -0.122*** | | -0.127*** |
| | | | (0.010) | | (0.009) |
| Lag_LS | 0.386*** | 0.305*** | | | |
| | (0.059) | (0.060) | | | |
| Lag_LS=2 | | | | 0.978*** | |
| | | | | (0.098) | |
| Lag_LS=3 | | | | 1.422*** | |
| | | | | (0.153) | |
| Lag_LS=4 | | | | 1.862*** | |
| | | | | (0.186) | |
| Lag_LS=5 | | | | 2.447*** | |
| | | | | (0.230) | |
| Age | 0.016 | -0.052*** | 0.003 | -0.042*** | 0.002 |
| | (0.036) | (0.007) | (0.005) | (0.007) | (0.005) |
| Age_sq | -0.106** | 0.038*** | 0.005 | 0.028*** | 0.006 |
| | (0.043) | (0.009) | (0.006) | (0.008) | (0.006) |
| Edu | 0.047** | -0.000 | 0.010*** | -0.002 | 0.010*** |
| | (0.019) | (0.003) | (0.002) | (0.003) | (0.002) |
| Marry | -0.200 | -0.066 | 0.149*** | -0.102** | 0.149*** |
| | (0.220) | (0.044) | (0.032) | (0.041) | (0.032) |
| Male | 0.710*** | 0.201*** | -0.087*** | 0.193*** | -0.086*** |
| | (0.175) | (0.026) | (0.019) | (0.025) | (0.019) |
| Family_Tie | -0.050* | -0.020*** | 0.005 | -0.019*** | 0.005 |
| | (0.028) | (0.007) | (0.006) | (0.007) | (0.006) |
| Urban | -0.490*** | -0.231*** | 0.232*** | -0.258*** | 0.231*** |
| | (0.123) | (0.034) | (0.023) | (0.031) | (0.023) |
| N | 36,680 | 36,680 | 36,680 | 36,680 | 36,680 |
| City Controls | YES | YES | YES | YES | YES |
| Fixed Effects | YES | YES | YES | YES | YES |

Note: Robust standard errors in parentheses. ***Significant at 1%; **Significant at 5%; *Significant at 10%. The full tables are in Tables 3.27 and 3.28. Fixed effects include individual and time effects. By combining with Table 3.7, the coefficient in columns (3) is statistically significant at 1% level, which means that the instrumental variables are significant in predicting the endogenous explanatory variables in the first stage. Moreover, for the Sanderson-Windmeijer multivariate F tests of excluded instruments is 104.690 ($p-value = 0.000$) the IV F-stat is 184.20, and Anderson LM statistic is 103.988 ($p-value = 0.000$), which indicate the IVs pass the underidentification test. Cragg-Donald Wald F statistic is equal to 104.692, and according to Stock and Yogo (2005), the 10% maximal size is 16.380, the 104.692 is greater than the critical value, therefore, the joint instrumental variable passes the weak identification test.

Table 3.8: The Marginal Effects of the Variables of Interest on Return Mobility Decisions

|  | (1) Return_Mig | | (2) Return_Mig | |
|---|---|---|---|---|
|  | b | se | b | se |
| Lag_LS | 0.074*** | 0.015 | | |
| Lag_LS=2 | | | 0.106*** | 0.005 |
| Lag_LS=3 | | | 0.206*** | 0.015 |
| Lag_LS=4 | | | 0.337*** | 0.030 |
| Lag_LS=5 | | | 0.538*** | 0.051 |
| Age | -0.013*** | 0.002 | -0.011*** | 0.002 |
| Age_sq | 0.009*** | 0.002 | 0.007*** | 0.002 |
| Edu | -0.000 | 0.001 | -0.001 | 0.001 |
| Marry | -0.016 | 0.011 | -0.026** | 0.010 |
| Male | 0.049*** | 0.006 | 0.049*** | 0.006 |
| Family_Tie | -0.005*** | 0.002 | -0.005*** | 0.002 |
| ln_Pop_density | 0.014** | 0.007 | 0.012* | 0.007 |
| Unemployrate | -0.001 | 0.007 | -0.001 | 0.007 |
| Industrialratio | 0.078*** | 0.015 | 0.065*** | 0.015 |
| Religion | -0.070*** | 0.025 | -0.066*** | 0.025 |
| Hospital | 0.070*** | 0.024 | 0.061*** | 0.023 |
| School | -0.036*** | 0.008 | -0.034*** | 0.008 |
| Urban | -0.056*** | 0.008 | -0.066*** | 0.008 |
| N | 36,680 | | 36,680 | |

Note: Robust standard errors in parentheses. ***Significant at 1%; **Significant at 5%; *Significant at 10%. In column (2), the marginal effects of Lag_LS are reported by treating the lagged LS as the ordered variable.

economic or social well-being. Schiele (2021) delineates the factors influencing life satisfaction among return migrants from two perspectives. First, the decision for immigrants to stay or return is shaped by their expectations and experiences in both the host and origin countries, with unmet expectations in the host country fostering a higher intention to return. Conversely, the decision to return despite high life satisfaction can be driven by the intrinsic value of cultural and emotional ties to the place of origin, which may surpass the benefits derived from life satisfaction in the host country. This indicates that the utility function of return migrants is significantly influenced by life satisfaction, suggesting that even content migrants might harbor a strong inclination to return due to personal or familial motivations.

In Table 3.9, column (1), the coefficient of Return_Mig is -0.059, indicating that the decision to return to migrate has a negative impact on life satisfaction, and this result is significant at the 10% significance level. In column (2), after using the instrumental variable method (CMP-IV), the coefficient of Return_Mig becomes -0.533, indicating that after considering the potential endogeneity problem, the negative impact of the return migration decision on life satisfaction is greater, and this result is significant at the 1% significance level. In column (3), the coefficient of Distance is -0.002, and this result is significant at the 1% significance level. This shows that the instrumental variable has a strong correlation with the endogenous variable in the first-stage results.

Overall, it seems that the return migration decision has a significant negative impact on life satisfaction, and the impact is more significant after considering the endogeneity problem. This result is verified in the marginal effect results. From the marginal effect point of view, after making the return migration decision, the probability of life satisfaction of 1 increased by 4.5 p.p., the probability of life satisfaction of 2 increased by 4.9 p.p., the probability of life satisfaction of 3 increased by 9.2 p.p., the probability of life satisfaction of 4 decreased by 3.2 p.p., and the probability of life satisfaction of 5 decreased by 15.4 p.p. From the above results, it can be seen that making a return migration decision (Return_Mig = 1) has a positive effect on the probability of low levels of life satisfaction (LS = 1, 2, 3) and a negative effect on the probability of high levels of life satisfaction (LS = 4, 5). In other words, the return migration decision increases the probability of low satisfaction levels while reducing the probability of high satisfaction levels. This suggests that the return migration decision may be associated with a decrease in life satisfaction.

Such results may be due to the disruption of personal living arrangements and social networks, and the re-adaptation process after returning home may be challenging, making it difficult for returnees to reintegrate into their home society and rebuild their social networks, resulting in lower overall life satisfaction (Nowok et al., 2013). On the other hand, studies comparing the life satisfaction of immigrants, left-behinds, and

Table 3.9: Estimates the Impact of Return Migration Decisions on Life Satisfaction

| VARIABLES | (1)<br>xtoprobit<br>LS | (2)<br>CMP-IV<br>LS | (3)<br>CMP-IV<br>Mig |
|---|---|---|---|
| Distance | | | -0.002*** |
| | | | (0.001) |
| Return_Mig | -0.059* | -0.533*** | |
| | (0.031) | (0.119) | |
| Age | -0.072*** | -0.069*** | -0.175*** |
| | (0.007) | (0.008) | (0.010) |
| Age_sq | 0.088*** | 0.078*** | 0.166*** |
| | (0.008) | (0.009) | (0.011) |
| Edu | -0.003 | -0.008*** | 0.003 |
| | (0.003) | (0.003) | (0.004) |
| Marry | 0.352*** | 0.197*** | 0.027 |
| | (0.038) | (0.035) | (0.049) |
| Male | -0.064** | 0.040* | 0.190*** |
| | (0.026) | (0.023) | (0.032) |
| Lag_LS | -0.009 | 0.323*** | -0.006 |
| | (0.013) | (0.011) | (0.015) |
| ln_Month_Income | -0.049*** | 0.010 | 0.021 |
| | (0.006) | (0.009) | (0.014) |
| ln_AVE_Social_Network | 0.045*** | 0.031*** | -0.004 |
| $N$ | 30,579 | 30,579 | 30,579 |
| City Controls | YES | YES | YES |
| Fixed Effects | YES | YES | YES |

Note: Robust standard errors in parentheses. ***Significant at 1%; **Significant at 5%; *Significant at 10%. The full tables are in Tables 3.29 and 3.30. Fixed effects include individual and time effects. The constant term is from the first stage of IV estimation. In Table 3.9, the coefficient is statistically significant at 1% level, which means that the instrumental variables are significant in predicting the endogenous explanatory variables in the first stage. Moreover, for the Sanderson-Windmeijer multivariate F test of excluded instruments is 33.73 ($p-value = 0.000$) the IV F-stat is 104.06, and Anderson LM statistic is 33.688 ($p-value = 0.000$), which indicate the IVs pass the underidentification test. Cragg-Donald Wald F statistic is equal to 33.732, and according to Stock and Yogo (2005), the 10% maximal size is 16.38, the number is greater than the critical value, therefore, the joint instrumental variable passes the weak identification test.

Table 3.10: The Marginal Effects of the Return Migration Decision on Life Satisfaction

|  | (1) | |
| --- | --- | --- |
|  | LS | |
|  | b | se |
| i.Return_Mig=1 | | |
| LS=1 | 0.045*** | 0.011 |
| LS=2 | 0.049*** | 0.011 |
| LS=3 | 0.092*** | 0.018 |
| LS=4 | -0.032*** | 0.006 |
| LS=5 | -0.154*** | 0.034 |
| $N$ | 30,579 | |

Note: Robust standard errors in parentheses. ***Significant at 1%; **Significant at 5%; *Significant at 10%. Here reports when $Mig = 1$, the marginal effects respect each level of life satisfaction. The full tables of the marginal effect of each level of life satisfaction are in Tables 3.40 to 3.44.

returnees have shown that returned immigrants are generally less satisfied with their lives than those who remain in the host country. This is because returnees lose out on the benefits and opportunities available abroad, such as better health care, education, and social services (Lapshyna and Düvell, 2015; Baykara-Krumme and Platt, 2018).

In the study of the impact of life satisfaction on migration decisions due to heterogeneity in educational attainment, the findings reveal that the effect of lagged life satisfaction on migration decisions varies significantly across different education levels. For individuals with low education (Educ = 0) and comparing with the baseline, an increase when Lag_LS = 2 significantly boosts migration likelihood. This group might experience a more immediate response to improved life satisfaction, seeking better opportunities promptly. However, Lag_LS = 4 and Lag_LS = 5 have a negative impact, possibly due to a developing sense of contentment and reduced motivation to migrate after initial improvements. For those with intermediate education (Educ = 1) and comparing with the baseline, only Lag_LS = 2 shows a significant positive effect, which means that when individuals are at a low level of life satisfaction and have intermediate education, they are more likely to make moving decisions. However, on other levels, even though they gradually turn negative, the results are indifference from zero. Lastly, individuals with high education (Educ = 2) and compared with the baseline, they experience a positive impact when La_LS = 2, but significant negative impacts when Lag_LS = 3, Lag_LS = 4, and Lag_LS = 5. Comparing the three groups of results, regardless of the level of education, when life satisfaction is low, people are more likely to make migration decisions. But when the education level is higher, on the overall level, the possibility of people making migration decisions is lower than that of people with low education level. The interaction between high life satisfaction and education level shows that more

Table 3.11: The Heterogeneous Effects of Education Level on Migration Decisions Through Life Satisfaction

| VARIABLES | (1)<br>CMP-IV<br>Mig | (2)<br>CMP-IV<br>Lag_LS |
|---|---|---|
| Commute_time | | -0.005*** |
| | | (0.001) |
| 1.Educ#c.Commute_time | | 0.003** |
| | | (0.001) |
| 2.Educ#c.Commute_time | | 0.001 |
| | | (0.002) |
| 2.Lag_LS | 0.904*** | |
| | (0.068) | |
| 3.Lag_LS | -0.074 | |
| | (0.067) | |
| 4.Lag_LS | -0.151** | |
| | (0.070) | |
| 5.Lag_LS | -0.189*** | |
| | (0.071) | |
| 1.Educ | -0.144 | -0.038 |
| | (0.097) | (0.032) |
| 2.Educ | 0.294 | 0.168*** |
| | (0.199) | (0.045) |
| 2.Lag_LS#1.Educ | 0.090 | |
| | (0.104) | |
| 2.Lag_LS#2.Educ | -0.511** | |
| | (0.212) | |
| 3.Lag_LS#1.Educ | 0.103 | |
| | (0.102) | |
| 3.Lag_LS#2.Educ | -0.419** | |
| | (0.206) | |
| 4.Lag_LS#1.Educ | 0.061 | |
| | (0.104) | |
| 4.Lag_LS#2.Educ | -0.399* | |
| | (0.206) | |
| 5.Lag_LS#1.Educ | 0.139 | |
| | (0.106) | |
| 5.Lag_LS#2.Educ | -0.449** | |
| | (0.214) | |
| Individual Controls | YES | |
| City Controls | YES | |
| N | 36,680 | 36,680 |
| Fixed Effects | YES | YES |

Note: Robust standard errors are in parentheses. ***Significant at 1%; **Significant at 5%; *Significant at 10%. The full tables are in Tables 3.31 and 3.32. Fixed effects include individual and time effects.

Table 3.12: The Marginal Effects of the Heterogeneous Effects of Education Level on Migration Decisions through Life Satisfaction

|  | (1) | | (2) | | (3) | |
|---|---|---|---|---|---|---|
|  | Mig (Educ=0) | | Mig (Educ=1) | | Mig (Educ=2) | |
|  | b | se | b | se | b | se |
| Lag_LS=2 | 0.287*** | 0.019 | 0.303*** | 0.020 | 0.129** | 0.062 |
| Lag_LS=3 | -0.019 | 0.017 | 0.007 | 0.018 | -0.130** | 0.059 |
| Lag_LS=4 | -0.037** | 0.018 | -0.020 | 0.019 | -0.142** | 0.059 |
| Lag_LS=5 | -0.045** | 0.018 | -0.011 | 0.020 | -0.160*** | 0.060 |
| Educ=1 | -0.012*** | 0.005 | -0.012*** | 0.005 | -0.012*** | 0.005 |
| Educ=2 | -0.031*** | 0.008 | -0.031*** | 0.008 | -0.031*** | 0.008 |
| Age | -0.037*** | 0.001 | -0.036*** | 0.001 | -0.034*** | 0.001 |
| Age_sq | 0.032*** | 0.001 | 0.031*** | 0.001 | 0.030*** | 0.001 |
| Marry | 0.003 | 0.008 | 0.003 | 0.007 | 0.003 | 0.007 |
| Male | 0.065*** | 0.004 | 0.063*** | 0.004 | 0.061*** | 0.004 |
| Family_Tie | -0.007*** | 0.001 | -0.006*** | 0.001 | -0.006*** | 0.001 |
| ln_Pop_density | -0.015*** | 0.003 | -0.015*** | 0.003 | -0.014*** | 0.003 |
| Unemployrate | -0.018*** | 0.006 | -0.017*** | 0.006 | -0.016*** | 0.006 |
| Industrialratio | 0.012** | 0.005 | 0.012** | 0.005 | 0.012** | 0.005 |
| Religion | -0.008 | 0.019 | -0.007 | 0.019 | -0.007 | 0.018 |
| Hospital | -0.027** | 0.011 | -0.026** | 0.011 | -0.025** | 0.010 |
| School | 0.018*** | 0.003 | 0.018*** | 0.003 | 0.017*** | 0.003 |
| Urban | -0.046*** | 0.005 | -0.044*** | 0.005 | -0.042*** | 0.004 |
| $N$ | 36,680 | | 36,680 | | 36,680 | |

Note: Robust standard errors in parentheses. ***Significant at 1%; **Significant at 5%; *Significant at 10%. Here reports when $Mig = 1$, the marginal effects respect each level of life satisfaction. The full tables of the marginal effect of each level of life satisfaction are in Tables 3.40 to 3.44.

educated individuals with high life satisfaction are less likely to migrate compared to their less satisfied counterparts (Chindarkar, 2014). These results underscore the importance of considering educational heterogeneity when examining the relationship between life satisfaction and migration, as different educational backgrounds may lead to varying motivations and responses to changes in life satisfaction. These findings are consistent with literature from Chindarkar (2014); Bartram (2015); Schiele (2021).

## 3.7 Policy Implications

The dual findings on the relationship between life satisfaction and migration decisions provide critical insights into migration dynamics and their implications for policy design. The results underscore the importance of addressing life satisfaction as a determinant and outcome of migration, while also considering moderating factors like education levels and the unique challenges of return migration.

Policymakers should prioritise initiatives to improve urban life satisfaction by addressing critical issues such as housing affordability, traffic congestion, and public services. Enhanced quality of life in urban centers can discourage outflows of dissatisfied individuals, promoting social stability and retaining a skilled workforce.

Local governments should address the challenges of return migration by developing tailored support programs for return migrants, including reintegration assistance, job placement services, and local infrastructure improvements. These measures can reduce the negative impacts of return migration and make smaller cities or rural areas more attractive. They also need to promote regional development to attract return migrants. Invest in smaller urban and rural areas to enhance infrastructure, healthcare, and education, while creating employment opportunities. By narrowing the life satisfaction gap between urban and non-urban areas, these regions can attract return migrants and contribute to regional economic growth.

The study reveals complex relationships between migration, life satisfaction, and moderating factors like education, emphasizing the need for nuanced policymaking. Support further research to deepen understanding of how life satisfaction influences migration decisions across demographic groups. Insights from such studies can inform evidence-based policies and targeted interventions.

## 3.8 Conclusion

This study investigates the relationship between life satisfaction and migration decisions, focusing on the Chinese labour force. The analysis reveals several key

insights that contribute to understanding how life satisfaction influences migration behaviour, and is consistent with the literature.

The main findings are in two parts. First, The results indicate that lagged life satisfaction is a robust predictor of migration decisions. As additional control variables for individual demographic characteristics are gradually included, the negative correlation between lagged life satisfaction and migration decisions remains significant at the 1% statistical level. Specifically, at the overall average level, an increase in lagged life satisfaction is associated with a 6.3 p.p. decrease in the probability of making a migration decision, confirming the hypothesis that higher life satisfaction reduces the likelihood of migration. However, when examining life satisfaction at different levels, individuals with a lagged life satisfaction score of 2 are 29.0 p.p. more likely to migrate compared with baseline, indicating that lower life satisfaction strongly motivates migration. The higher mobility observed among individuals with LS score 2 compared to LS score 1 can be explained by differences in capabilities rather than aspirations; those with LS=2 likely have slightly better resources or support systems to act on their desire to move, aligning with Sen's emphasis on the interplay between aspirations and real opportunities. Conversely, individuals with lagged life satisfaction scores of 4 or 5 are less likely to migrate by 3.7 and 3.8 p.p., respectively compared with baseline. These findings align with the literature suggesting that individuals with higher life satisfaction are less inclined to migrate, preferring stability over change. Secondly, Migration decisions positively influence life satisfaction. The conditional marginal effects show that migrants are less likely to report lower life satisfaction and more likely to report higher life satisfaction compared to non-migrants. Migration is associated with a 9.5 p.p. increase in the probability of reporting the highest level of life satisfaction. This suggests that migration decisions are linked to an overall increase in life satisfaction, supporting the notion that migration can be a strategy for enhancing well-being.

In the additional analysis, the results find the following two things. The decision to leave big cities (return migration) is examined separately. The findings indicate that individuals with higher past life satisfaction are significantly more likely to return and migrate. However, return migration has a negative impact on life satisfaction, especially after addressing potential endogeneity issues. The probability of reporting lower life satisfaction levels increases, while the probability of reporting higher life satisfaction levels decreases after return migration. This may be due to the challenges of reintegration and the loss of benefits associated with residing in larger cities. The impact of life satisfaction on migration decisions varies across different education levels. Individuals with low education levels are more likely to migrate when their life satisfaction is low, while those with higher education levels

exhibit a lower propensity to migrate as their life satisfaction increases. This suggests that education moderates the relationship between life satisfaction and migration decisions, with more educated individuals possibly having better-coping mechanisms or opportunities in their current locations.

These findings have several implications for policymakers and researchers. Policymakers should consider the role of life satisfaction in shaping migration decisions. Enhancing the quality of life in urban areas can reduce the outflow of dissatisfied individuals and improve overall social stability. Conversely, improving life satisfaction in rural or smaller urban areas could attract return migrants, contributing to balanced regional development.

In analysing the relationship between migration mobility and life satisfaction, key limitations include the use of a single-item scale to measure life satisfaction, which may not fully capture its complexity. Additionally, the biennial data collection may miss shorter-term fluctuations. The use of commute time and overtime work hours as instrumental variables may not fully address endogeneity concerns, potentially limiting the accuracy of the findings. Furthermore, the small sample size, especially within migrant sub-groups, restricts the generalisability of the results. Addressing these issues in future research could enhance the validity and reliability of the conclusions.

For future research, the study highlights the need to consider the heterogeneity of the migrant population, particularly in terms of education and past life satisfaction levels. Further exploration of the mechanisms through which life satisfaction influences migration decisions can provide deeper insights into the complex interplay between individual well-being and mobility.

In conclusion, this study underscores the importance of life satisfaction as a significant determinant of migration decisions. Both higher and lower levels of life satisfaction can drive migration behavior, with significant implications for individual well-being and policy design.

# 3.A  Appendix

## 3.A.1  Additional Tables

The results in Table 3.13 indicate no statistically significant differences. The mean commuting time for migrants shows almost no change between 2012 and 2014, with a difference of -0.001 and a p-value of 0.998, indicating no significance. For non-migrants, the mean commuting time difference is 0.008, also statistically insignificant with a p-value of 0.983. The difference in mean commuting time between migrants and non-migrants in 2012 is -0.435 (p-value = 0.465), and in 2014 is -0.444 (p-value = 0.454), both of which are not statistically significant. These results suggest that the difference in commuting time between 2012 and 2014 is not significant for either migrants or non-migrants, indicating that migration decisions did not significantly impact commuting time during this period.

Table 3.13: Differences test in commuting times between migrants and non-migrants in 2012 and 2014

Two-sample t test with equal variances

|  | obs1 | obs2 | diff | p-value |
|---|---|---|---|---|
| Commute_time in 2014 (to Migrants) | 857 | 858 | -0.002 | 0.999 |
| Commute_time in 2014 (to Non-Migrants) | 3,316 | 3,279 | 0.008 | 0.983 |
| Commute_time in 2012 (between Migrants and Non-Migrants) | 3,316 | 857 | -0.435 | 0.465 |
| Commute_time (between Migrants and Non-Migrants) | 3,279 | 858 | -0.445 | 0.454 |

Note: Null Hypothesis (H0): The null hypothesis typically assumes that there is no significant difference between the means being compared.

Table 3.14 presents the results of testing the difference in commuting time between migrants and non-migrants in 2014 and 2016. The mean difference in commuting time for migrants between 2014 and 2016 is -0.477, which is not statistically significant (p-value = 0.517). Similarly, the mean difference for non-migrants during the same period is 0.007, also statistically insignificant (p-value = 0.986). The mean difference in commuting time between migrants and non-migrants in 2014 is -0.621 (p-value = 0.287), not statistically significant. However, in 2016, the mean difference is -1.104, and this difference is statistically significant (p-value = 0.043), indicating that migrants have significantly longer commuting times than non-migrants in 2016. These findings suggest that, while there is no significant change in commuting time between 2014 and 2016, migrants had significantly longer commuting times than non-migrants in 2016.

Table 3.14: Differences test in commuting times between migrants and non-migrants in 2014 and 2016

Two-sample t test with equal variances

|  | obs1 | obs2 | diff | p-value |
|---|---|---|---|---|
| Commute_time in 2016 (to Migrants) | 902 | 1,039 | -0.477 | 0.517 |
| Commute_time in 2016 (to Non-Migrants) | 3,235 | 3,217 | 0.007 | 0.986 |
| Commute_time in 2014 (between Migrants and Non-Migrants) | 3,235 | 902 | -0.621 | 0.287 |
| Commute_time in 2016 (between Migrants and Non-Migrants) | 3,217 | 1,039 | -1.105 | 0.043 |

Note: Null Hypothesis (H0): The null hypothesis typically assumes that there is no significant difference between the means being compared.

Table 3.15 shows the results of testing the difference in commuting time between migrants and non-migrants in 2016 and 2018. The results show that there is no significant difference in the commuting time of migrants in 2016 and 2018 (the difference is 0.513, p = 0.449). This indicates that there is no significant change in the commuting time of migrants between the two years. There is also no significant change in the commuting time of non-migrants between the two years (the difference is -0.151, p = 0.692). In 2016, the commuting time of migrants is slightly longer than that of non-migrants, but the difference is not significant (the difference is -0.774, p = 0.149). In 2018, there is no significant difference in the commuting time of migrants and non-migrants (the difference is -0.110, p = 0.839). From the comparison of the commuting time of migrants and non-migrants in 2016 and 2018, the migration decision did not lead to a significant change in commuting time. The commuting time of migrants at the two-time points (2016 and 2018) did not decrease significantly. On the contrary, the commuting time of migrants is slightly longer in some years. This suggests that the motivations for migration decisions may be unrelated to commuting time.

Table 3.15: Differences test in commuting times between migrants and non-migrants in 2016 and 2018

Two-sample t test with equal variances

|  | obs1 | obs2 | diff | p-value |
|---|---|---|---|---|
| Commute_time in 2018 (to Migrants) | 1,088 | 1,092 | 0.513 | 0.450 |
| Commute_time in 2018 (to Non-Migrants) | 3,168 | 3,122 | -.0151 | 0.692 |
| Commute_time in 2016 (between Migrants and Non-Migrants) | 3,168 | 1,088 | -0.774 | 0.149 |
| Commute_time in 2018 (between Migrants and Non-Migrants) | 3,122 | 1,092 | -0.110 | 0.839 |

Note: Null Hypothesis (H0): The null hypothesis typically assumes that there is no significant difference between the means being compared.

Table 3.16 shows the results of testing the difference in commuting time between migrants and non-migrants in 2018 and 2020. The results show a significant difference in the commuting time of migrants in 2018 and 2020 (the difference is -1.433, p = 0.032). The commuting time of migrants in 2020 increased significantly. The commuting time of non-migrants in 2018 and 2020 does not change considerably between the two years (the difference is -0.009, p = 0.981). There is no significant difference in the commuting time between migrants and non-migrants in 2018 (the difference is -0.039, p = 0.942). There is a significant difference in the commuting time between migrants and non-migrants in 2020, with migrants commuting significantly longer than non-migrants (the difference is -1.463, p = 0.0062). Migrants' commuting time in 2020 is significantly longer than in 2018. This suggests that migration decisions may have led to an increase in commuting time, rather than a decrease. This suggests that the commuting burden of migrants is increasing, which in turn suggests that migrants' migration decisions may not be driven by a desire to reduce commuting time.

Table 3.16: Differences test in commuting times between migrants and non-migrants in 2018 and 2020

Two-sample t test with equal variances

|  | obs1 | obs2 | diff | p-value |
|---|---|---|---|---|
| Commute_time in 2020 (to Migrants) | 1,113 | 1,208 | -1.433 | 0.032 |
| Commute_time in 2020 (to Non-Migrants) | 3,101 | 2,942 | -0.009 | 0.982 |
| Commute_time in 2018 (between Migrants and Non-Migrants) | 3,101 | 1,113 | -0.039 | 0.943 |
| Commute_time in 2020 (between Migrants and Non-Migrants) | 2,942 | 1,208 | -1.463 | 0.006 |

Note: Null Hypothesis (H0): The null hypothesis typically assumes that there is no significant difference between the means being compared.

Table 3.17 shows the results of testing the difference in overtime working hours between migrants and non-migrants in 2012 and 2014. The results showe that in 2012, the average overtime hours of migrants were significantly higher than those of non-migrants (the difference is 1.731, and the p-value is 0.000). In 2014, the average overtime hours of migrants is significantly higher than those of non-migrants (the difference is 1.413, and the p-value is 0.000). Comparing the overtime hours of migrants and non-migrants in 2012 and 2014, there is no significant change in the overtime hours of migrants in 2012 and 2014 (the difference is 0.000, and the p-value is 1.000). The overtime hours of non-migrants in 2012 and 2014 do not change significantly (the difference is -0.317, and the p-value is 1.000). This indirectly means that the migration decision of migrants is not necessarily affected by the length of overtime hours.

Table 3.17: Differences test in overtime work between migrants and non-migrants in 2012 and 2014

Two-sample t test with equal variances

|  | obs1 | obs2 | diff | p-value |
|---|---|---|---|---|
| overtime_work in 2014 (to Migrants) | 1,707 | 1,708 | 0.000 | 1.000 |
| overtime_work in 2014 (to Non-Migrants) | 7,992 | 7,681 | -0.318 | 0.142 |
| overtime_work in 2012 (between Migrants and Non-Migrants) | 7,992 | 1,707 | -1.730 | 0.000 |
| overtime_work in 2014 (between Migrants and Non-Migrants) | 7,681 | 1,708 | -1.413 | 0.000 |

Note: Null Hypothesis (H0): The null hypothesis typically assumes that there is no significant difference between the means being compared.

Table 3.18 shows the results of testing the difference in overtime working hours between migrants and non-migrants in 2014 and 2016. The results show that the mean difference in overtime hours between migrants in 2014 and 2016 is -0.061, which is not statistically significant (p-value is 0.896). This indicates that there is no significant change in the overtime hours of migrants in these two years. The mean difference in overtime hours between non-migrants in 2014 and 2016 is -0.004, which is not statistically significant (p-value is 0.985). This indicates that there is no significant change in the overtime hours of non-migrants in these two years. The mean difference in overtime hours between migrants and non-migrants in 2014 is -1.620, which is statistically significant (p-value is 0.000). This indicates that in 2014, migrants worked significantly more overtime than non-migrants. The mean difference in overtime hours between migrants and non-migrants in 2016 is -1.677, which is statistically significant (p-value is 0.000). This indicates that in 2016, migrants worked significantly more overtime than non-migrants. These indirectly suggest that migration decisions are not affected by overtime hours.

Table 3.18: Differences test in overtime work between migrants and non-migrants in 2014 and 2016

Two-sample t test with equal variances

|  | obs1 | obs2 | diff | p-value |
|---|---|---|---|---|
| overtime_work in 2016 (to Migrants) | 1,760 | 1,829 | -0.061 | 0.896 |
| overtime_work in 2016 (to Non-Migrants) | 7,629 | 7,348 | -0.004 | 0.986 |
| overtime_work in 2014 (between Migrants and Non-Migrants) | 7,629 | 1,760 | -1.620 | 0.000 |
| overtime_work in 2016 (between Migrants and Non-Migrants) | 7,348 | 1,829 | -1.677 | 0.000 |

Note: Null Hypothesis (H0): The null hypothesis typically assumes that there is no significant difference between the means being compared.

Table 3.19 shows the results of testing the difference in overtime working hours between migrants and non-migrants in 2016 and 2018. The results show that the mean difference in overtime hours between migrants in 2016 and 2018 is -1.018, which is statistically significant (p-value is 0.022). This indicates that migrants worked significantly more overtime in 2018 than in 2016. The mean difference in overtime hours between non-migrants in 2016 and 2018 is -1.223, which is statistically significant (p-value is 0.000). This indicates that non-migrants worked significantly more overtime in 2018 than in 2016. The mean difference in overtime hours between migrants and non-migrants in 2016 is -1.845, which is statistically significant (p-value is 0.000). This indicates that migrants worked significantly more overtime than non-migrants in 2016. The mean difference in overtime hours between migrants and non-migrants in 2018 is -1.639, which is statistically significant (p-value is 0.000). This indicates that migrants worked significantly more overtime than non-migrants in 2018. These evidences indirectly suggest that overtime hours do not necessarily have a significant impact on migration decisions.

Table 3.19: Differences test in overtime work between migrants and non-migrants in 2016 and 2018

Two-sample t test with equal variances

|  | obs1 | obs2 | diff | p-value |
|---|---|---|---|---|
| overtime_work in 2018 (to Migrants) | 1,926 | 2,020 | -1.018 | 0.022 |
| overtime_work in 2018 (to Non-Migrants) | 7,251 | 7,153 | -1.223 | 0.000 |
| overtime_work in 2016 (between Migrants and Non-Migrants) | 7,251 | 1,926 | -1.845 | 0.000 |
| overtime_work in 2018 (between Migrants and Non-Migrants) | 7,153 | 2,020 | -1.639 | 0.000 |

Note: Null Hypothesis (H0): The null hypothesis typically assumes that there is no significant difference between the means being compared.

Table 3.20 shows the results of testing the difference in overtime working hours between migrants and non-migrants in 2016 and 2018. The results show that the mean difference in overtime hours between migrants in 2018 and 2020 is -0.701, but the difference is not statistically significant (p-value is 0.105). The mean difference in overtime hours between non-migrants in 2018 and 2020 is -0.032, which is not statistically significant (p-value is 0.890). The mean difference in overtime hours between migrants and non-migrants in 2018 is -1.990, which is statistically significant (p-value is 0.000). The mean difference in overtime hours between migrants and non-migrants in 2020 is -2.659, which is statistically significant (p-value is 0.000). These evidences may suggest that overtime hours do not necessarily have a significant impact on migration decisions.

Table 3.20: Differences test in overtime work between migrants and non-migrants in 2018 and 2020

Two-sample t test with equal variances

|  | obs1 | obs2 | diff | p-value |
| --- | --- | --- | --- | --- |
| overtime_work in 2020 (to Migrants) | 2,030 | 2,081 | -0.701 | 0.105 |
| overtime_work in 2020 (to Non-Migrants) | 7,143 | 6,601 | -0.032 | 0.890 |
| overtime_work in 2018 (between Migrants and Non-Migrants) | 7,143 | 2,030 | -1.990 | 0.000 |
| overtime_work in 2020 (between Migrants and Non-Migrants) | 6,601 | 2,081 | -2.659 | 0.000 |

Note: Null Hypothesis (H0): The null hypothesis typically assumes that there is no significant difference between the means being compared.

Table 3.21: Estimates Equation 1 using single equation

| VARIABLES | (1)<br>xtprobit<br>Mig | (2)<br>xtprobit<br>Mig | (3)<br>xtprobit<br>Mig | (4)<br>xtprobit<br>Mig | (5)<br>CMP-IV<br>Mig | (6)<br>CMP-IV<br>Lag_LS |
|---|---|---|---|---|---|---|
| Overtime_work | | | | | | -0.002** |
| | | | | | | (0.001) |
| Commute_time | | | | | | -0.003*** |
| | | | | | | (0.001) |
| Lag_LS | -0.277*** | -0.322*** | -0.317*** | -0.375*** | -0.250*** | |
| | (0.028) | (0.047) | (0.030) | (0.085) | (0.010) | |
| Age | | -0.806** | -0.720*** | -0.875*** | -0.156*** | -0.017*** |
| | | (0.379) | (0.059) | (0.075) | (0.005) | (0.006) |
| Age_sq | | 0.609*** | 0.575*** | 0.658*** | 0.139*** | 0.028*** |
| | | (0.179) | (0.077) | (0.091) | (0.006) | (0.007) |
| Edu | | | 0.015 | 0.045 | -0.009*** | 0.009*** |
| | | | (0.049) | (0.038) | (0.002) | (0.002) |
| Marry | | | | 0.197 | 0.021 | 0.201*** |
| | | | | (0.356) | (0.030) | (0.032) |
| Male | -0.323 | 1.466 | 1.079** | 1.256*** | 0.280*** | -0.044** |
| | (0.427) | (1.162) | (0.535) | (0.309) | (0.017) | (0.018) |
| Family_Tie | -0.056** | -0.048* | -0.052** | -0.051 | -0.026*** | 0.009* |
| | (0.025) | (0.029) | (0.025) | (0.055) | (0.005) | (0.005) |
| Urban | -0.369*** | -0.352*** | -0.341*** | -0.279 | -0.190*** | 0.114*** |
| | (0.080) | (0.118) | (0.089) | (0.230) | (0.018) | (0.021) |
| ln_Pop_density | 0.083 | -0.204 | -0.172 | -0.114 | -0.070*** | 0.058*** |
| | (0.171) | (0.220) | (0.252) | (0.188) | (0.011) | (0.011) |
| Unemployrate | -0.232** | -0.190* | -0.175 | -0.183 | -0.070*** | 0.053** |
| | (0.096) | (0.115) | (0.109) | (0.314) | (0.026) | (0.023) |
| Industrialratio | 0.187 | 0.213 | 0.209 | 0.264 | 0.047** | -0.030 |
| | (0.123) | (0.158) | (0.140) | (0.252) | (0.019) | (0.021) |
| Religion | -0.948*** | -0.689* | -0.702** | -0.436 | 0.015 | 0.284*** |
| | (0.299) | (0.360) | (0.319) | (0.929) | (0.078) | (0.085) |
| Hospital | -0.668*** | -0.619** | -0.677*** | -0.762 | -0.166*** | 0.067 |
| | (0.220) | (0.272) | (0.225) | (0.475) | (0.045) | (0.053) |
| School | -0.238 | -0.119 | -0.146 | -0.078 | 0.094*** | -0.010 |
| | (0.210) | (0.263) | (0.235) | (0.223) | (0.014) | (0.014) |

Note: Robust standard errors in parentheses. ***Significant at 1%; **Significant at 5%; *Significant at 10%.

Table 3.22: Continued Table 3.2

| VARIABLES | (1) xtprobit Mig | (2) xtprobit Mig | (3) xtprobit Mig | (4) xtprobit Mig | (5) CMP-IV Mig | (6) CMP-IV Lag_LS |
|---|---|---|---|---|---|---|
| Constant | -8.294*** | 12.497*** | 8.683*** | 13.672*** | 3.965*** | |
| | (1.644) | (4.715) | (2.073) | (1.851) | (0.135) | |
| cut_2_1 | | | | | | -1.343*** |
| | | | | | | (0.148) |
| cut_2_2 | | | | | | -0.392*** |
| | | | | | | (0.147) |
| cut_2_3 | | | | | | 0.600*** |
| | | | | | | (0.148) |
| cut_2_4 | | | | | | 1.454*** |
| | | | | | | (0.148) |
| $N$ | 37,094 | 37,094 | 36,818 | 36,680 | 36,680 | 36,680 |
| Fixed Effects | YES | YES | YES | YES | YES | YES |

Note: Robust standard errors in parentheses. ***Significant at 1%; **Significant at 5%; *Significant at 10%. Fixed effects include individual and time effects. The cut points indicate the thresholds for different categories of choice observed in our sample. We report the OLS results in Table 3.33. The intuition is that preliminary results based on the span of data type methods can still show us the signs associated with our coefficients. I report the IV specifications in Tables 3.45 and 3.46, by combining with Tables 3.21 and 3.22, the coefficient is statistically significant at 1% level, which means that the instrumental variables are significant in predicting the endogenous explanatory variables in the first stage. Moreover, for the Sanderson-Windmeijer multivariate F test of excluded instruments is 15.35 ($p-value = 0.000$) the IV F-stat is 110.51, and Anderson LM statistic is 30.676 ($p-value = 0.000$), which indicate the IVs pass the underidentification test. Cragg-Donald Wald F statistic is equal to 15.348, and according to Hausman et al. (2005); Stock and Yogo (2005), the 10% maximal size is 8.68, the 15.348 is greater than Stock-Yogo weak ID test critical value, therefore, the joint instrumental variable passes weak identification test. Lastly, Hansen's test fails to reject the null that the instruments are valid ($J-test = 0.069$, $p-value = 0.793$), which means our joint instrumental variable is not over-identified.

In Tables 3.45 and 3.46, I report the IV specifications. Columns 1,3 and 5 are the first stage of 2SLS results, and columns 2, 4, and 6 are the second stage results. The coefficients in the first stage are statistically significant at 1% level, which means that the instrumental variables are significant in predicting the endogenous explanatory variables in the first stage. Columns 1 and 2 report when I only use overtime work as a single instrumental variable; columns 3 and 4 report when I only use commute time as a single instrumental variable; the last two columns report when I jointly use overtime work and commute time as instruments. For the Sanderson-Windmeijer multivariate F tests of excluded instruments are 14.548, 22.740, and 15.348, all with respect to $p-value = 0.000$), which rejects the null that the particular endogenous regressor in question is unidentified. The IV F-stats are 235.22, 105.10, and 110.51, and Anderson LM statistics are 14.548, 22.732, and 30.676, all respect to $p-value = 0.000$, which indicates the IVs pass the underidentification test. Cragg-Donald Wald F statistics are equal to 14.545, 22.736, and 15.348; however, when we only use one

Table 3.23: Estimates Equation 1 using single equation by treating the lagged LS as ordered variable

| VARIABLES | (1) CMP-IV Mig | (2) CMP-IV Lag_LS |
|---|---|---|
| Overtime_work | | -0.002** |
| | | (0.001) |
| Commute_time | | -0.003*** |
| | | (0.001) |
| Lag_LS=2 | 0.928*** | |
| | (0.050) | |
| Lag_LS=3 | -0.060 | |
| | (0.050) | |
| Lag_LS=4 | -0.158*** | |
| | (0.053) | |
| Lag_LS=5 | -0.162*** | |
| | (0.057) | |
| Age | -0.149*** | -0.017*** |
| | (0.005) | (0.006) |
| Age_sq | 0.132*** | 0.028*** |
| | (0.006) | (0.007) |
| Edu | -0.007*** | 0.009*** |
| | (0.002) | (0.002) |
| Marry | 0.018 | 0.201*** |
| | (0.031) | (0.032) |
| Male | 0.266*** | -0.044** |
| | (0.017) | (0.018) |
| Family_Tie | -0.027*** | 0.009* |
| | (0.005) | (0.005) |
| Urban | -0.181*** | 0.114*** |
| | (0.019) | (0.021) |
| ln_Pop_density | -0.063*** | 0.058*** |
| | (0.011) | (0.011) |
| Unemployrate | -0.067** | 0.053** |
| | (0.026) | (0.023) |
| Industrialratio | 0.055*** | -0.030 |
| | (0.020) | (0.021) |
| Religion | -0.001 | 0.283*** |
| | (0.079) | (0.085) |
| Hospital | -0.162*** | 0.067 |
| | (0.046) | (0.053) |
| School | 0.082*** | -0.010 |
| | (0.014) | (0.014) |

Note: Robust standard errors in parentheses. ***Significant at 1%; **Significant at 5%; *Significant at 10%.

Table 3.24: Continued Table 3.23

| VARIABLES | (1) CMP-IV Mig | (2) CMP-IV Lag_LS |
|---|---|---|
| Constant | 2.743*** | |
| | (0.143) | |
| cut_2_1 | | -1.345*** |
| | | (0.148) |
| cut_2_2 | | -0.394*** |
| | | (0.147) |
| cut_2_3 | | 0.600*** |
| | | (0.148) |
| cut_2_4 | | 1.454*** |
| | | (0.148) |
| N | 36,680 | 36,680 |
| Fixed Effects | YES | YES |

Note: Robust standard errors in parentheses. ***Significant at 1%; **Significant at 5%; *Significant at 10%. Fixed effects include individual and time effects. The cut points indicate the thresholds for different categories of choice observed in our sample.

instrumental variable, according to Hausman et al. (2005); Stock and Yogo (2005), the 10% maximal size of the Stock-Yogo weak ID test critical value is 16.38. This indicates, overtime work, as IV falls to weak identification, which arises when the excluded instruments are correlated with the endogenous regressors, but only weakly. Therefore, I employ the second IV, commute time, in 2SLS estimation, and it applies all the 2SLS specifications. Lastly, I jointly employ overtime work and commute time as IV, and the statistical evidence is satisfied with the above specifications and Hansen's test fails to reject the null that the instruments are valid ($J - test = 0.069$, $p - value = 0.793$), which means our joint instrumental variable is not over-identified.

Table 3.25: Estimates Equation 2 using single equation

| VARIABLES | (1) LS | (2) LS | (3) LS | (4) LS | (5) CMP-IV LS | (6) CMP-IV Mig |
|---|---|---|---|---|---|---|
| Distance | | | | | | -0.002*** |
| | | | | | | (0.000) |
| Mig | 0.057** | 0.037* | 0.033* | 0.027* | 0.308*** | |
| | (0.023) | (0.025) | (0.025) | (0.025) | (0.048) | |
| Age | | -0.044*** | -0.045*** | -0.065*** | -0.032*** | -0.168*** |
| | | (0.005) | (0.005) | (0.006) | (0.005) | (0.006) |
| Age_sq | | 0.057*** | 0.056*** | 0.078*** | 0.042*** | 0.147*** |
| | | (0.006) | (0.006) | (0.006) | (0.005) | (0.007) |
| Edu | | | -0.011*** | -0.011*** | -0.009*** | -0.005** |
| | | | (0.002) | (0.002) | (0.002) | (0.002) |
| Marry | | | | 0.311*** | 0.242*** | -0.050 |
| | | | | (0.031) | (0.022) | (0.031) |
| Male | -0.044** | -0.053*** | -0.038** | -0.032* | -0.032** | 0.324*** |
| | (0.018) | (0.018) | (0.019) | (0.019) | (0.013) | (0.019) |
| Lag_LS | 0.150*** | 0.146*** | 0.151*** | 0.149*** | 0.336*** | -0.114*** |
| | (0.009) | (0.009) | (0.010) | (0.012) | (0.006) | (0.009) |
| ln_Month _Income | 0.007 | 0.019*** | 0.021*** | 0.019*** | 0.017*** | -0.003 |
| | (0.005) | (0.006) | (0.006) | (0.006) | (0.005) | (0.008) |
| ln_AVE_ Social_Network | -0.003 | 0.004 | 0.012* | 0.011* | 0.013** | -0.026*** |
| | (0.007) | (0.007) | (0.007) | (0.007) | (0.006) | (0.009) |
| ln_CO2 | 0.008 | 0.009 | 0.015 | 0.015 | 0.014 | -0.071*** |
| | (0.018) | (0.018) | (0.018) | (0.017) | (0.012) | (0.019) |
| ln_SO2 | 0.020** | 0.020** | 0.022*** | 0.021** | 0.011 | 0.036*** |
| | (0.008) | (0.008) | (0.008) | (0.008) | (0.007) | (0.009) |
| AQI | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | -0.003 |
| | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.002) |
| PM2_5 | -0.001 | -0.000 | -0.000 | -0.000 | -0.000 | -0.000 |
| | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) |
| PM10 | 0.000 | 0.000 | 0.000 | 0.000 | -0.000 | 0.002*** |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.001) |
| Greenrate | -0.001* | -0.001* | -0.001* | -0.001* | -0.001** | -0.000 |
| | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) |

Note: Robust standard errors in parentheses. ***Significant at 1%; **Significant at 5%; *Significant at 10%.

Table 3.26: Continued Table 3.25

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| VARIABLES | LS | LS | LS | LS | CMP-IV LS | CMP-IV Mig |
| Constant | | | | | | 1.780*** |
| | | | | | | (0.198) |
| cut_1_1 | -0.923*** | -1.560*** | -1.621*** | -1.841*** | -0.443*** | |
| | (0.124) | (0.170) | (0.171) | (0.173) | (0.146) | |
| cut_1_2 | -0.254** | -0.890*** | -0.952*** | -1.171*** | 0.148 | |
| | (0.123) | (0.169) | (0.170) | (0.172) | (0.145) | |
| cut_1_3 | 1.103*** | 0.471*** | 0.411** | 0.196 | 1.342*** | |
| | (0.122) | (0.167) | (0.169) | (0.170) | (0.144) | |
| cut_1_4 | 2.092*** | 1.463*** | 1.404*** | 1.189*** | 2.207*** | |
| | (0.122) | (0.167) | (0.168) | (0.169) | (0.143) | |
| N | 30,847 | 30,847 | 30,579 | 30,579 | 30,579 | 30,579 |
| Fixed Effects | YES | YES | YES | YES | YES | YES |

Note: Robust standard errors in parentheses. ***Significant at 1%; **Significant at 5%; *Significant at 10%. The cut points indicate the thresholds for different categories of choice observed in our sample. The constant term is from the first stage of IV estimation. We report the OLS results in Table 3.34. The intuition is that preliminary results based on the span of data type methods can still show us the signs associated with our coefficients. I report the IV specifications in Tables 3.47 and 3.48, by combining with Tables 3.25 and 3.26, the coefficient is statistically significant at 1% level, which means that the instrumental variables are significant in predicting the endogenous explanatory variables in the first stage. Moreover, for the Sanderson-Windmeijer multivariate F test of excluded instruments is 83.250 ($p-value = 0.000$) the IV F-stat is 226.88, and Anderson LM statistic is 83.079 ($p-value = 0.000$), which indicate the IVs pass the underidentification test. Cragg-Donald Wald F statistic is equal to 83.252, and according to Stock and Yogo (2005), the 10% maximal size is 8.68, the 15.348 is greater than the critical value, therefore, the joint instrumental variable passes the weak identification test.

Table 3.27: the Estimation of Lagged Life Satisfaction on Returning Migrations

| VARIABLES | (1) Return_Mig | (2) CMP-IV Return_Mig | (3) CMP-IV Lag_LS | (4) CMP-IV Return_Mig | (5) CMP-IV Lag_LS |
|---|---|---|---|---|---|
| Commute_time | | | -0.122*** | | -0.127*** |
| | | | (0.010) | | (0.009) |
| Lag_LS | 0.386*** | 0.305*** | | | |
| | (0.059) | (0.060) | | | |
| 2.Lag_LS | | | | 0.978*** | |
| | | | | (0.098) | |
| 3.Lag_LS | | | | 1.422*** | |
| | | | | (0.153) | |
| 4.Lag_LS | | | | 1.862*** | |
| | | | | (0.186) | |
| 5.Lag_LS | | | | 2.447*** | |
| | | | | (0.230) | |
| Age | 0.016 | -0.052*** | 0.003 | -0.042*** | 0.002 |
| | (0.036) | (0.007) | (0.005) | (0.007) | (0.005) |
| Age_sq | -0.106** | 0.038*** | 0.005 | 0.028*** | 0.006 |
| | (0.043) | (0.009) | (0.006) | (0.008) | (0.006) |
| Edu | 0.047** | -0.000 | 0.010*** | -0.002 | 0.010*** |
| | (0.019) | (0.003) | (0.002) | (0.003) | (0.002) |
| Marry | -0.200 | -0.066 | 0.149*** | -0.102** | 0.149*** |
| | (0.220) | (0.044) | (0.032) | (0.041) | (0.032) |
| Male | 0.710*** | 0.201*** | -0.087*** | 0.193*** | -0.086*** |
| | (0.175) | (0.026) | (0.019) | (0.025) | (0.019) |
| Family_Tie | -0.050* | -0.020*** | 0.005 | -0.019*** | 0.005 |
| | (0.028) | (0.007) | (0.006) | (0.007) | (0.006) |
| Urban | -0.490*** | -0.231*** | 0.232*** | -0.258*** | 0.231*** |
| | (0.123) | (0.034) | (0.023) | (0.031) | (0.023) |
| ln_Pop_density | 0.711*** | 0.057** | -0.007 | 0.046* | -0.008 |
| | (0.177) | (0.028) | (0.019) | (0.026) | (0.019) |
| Unemployrate | -0.121 | -0.004 | 0.034 | -0.002 | 0.038 |
| | (0.118) | (0.027) | (0.024) | (0.026) | (0.023) |
| Industrialratio | 0.859*** | 0.318*** | 0.199*** | 0.254*** | 0.196*** |
| | (0.234) | (0.064) | (0.042) | (0.061) | (0.042) |
| Religion | -0.582 | -0.286*** | 0.064 | -0.261*** | 0.065 |
| | (0.381) | (0.104) | (0.076) | (0.097) | (0.076) |
| Hospital | 2.176*** | 0.284*** | 0.134* | 0.238*** | 0.136* |
| | (0.534) | (0.098) | (0.074) | (0.092) | (0.074) |
| School | -1.094*** | -0.148*** | 0.051** | -0.133*** | 0.051** |
| | (0.207) | (0.032) | (0.022) | (0.030) | (0.022) |

Note: Robust standard errors in parentheses. ***Significant at 1%; **Significant at 5%; *Significant at 10%.

Table 3.28: Continued Table 3.27

|  | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| VARIABLES | Return_Mig | CMP-IV Return_Mig | CMP-IV Lag_LS | CMP-IV Return_Mig | CMP-IV Lag_LS |
| Constant | -13.611*** | -2.283*** | | -2.368*** | |
|  | (1.905) | (0.231) | | (0.223) | |
| cut_2_1 | | | 1.385*** | | 1.422*** |
|  | | | (0.203) | | (0.197) |
| cut_2_2 | | | 3.079*** | | 3.112*** |
|  | | | (0.203) | | (0.198) |
| cut_2_3 | | | 3.785*** | | 3.816*** |
|  | | | (0.204) | | (0.199) |
| cut_2_4 | | | 4.498*** | | 4.531*** |
|  | | | (0.205) | | (0.199) |
| N | 36,680 | 36,680 | 36,680 | 36,680 | 36,680 |
| Fixed Effects | YES | YES | YES | YES | YES |

Note: Robust standard errors in parentheses. ***Significant at 1%; **Significant at 5%; *Significant at 10%. The cut points indicate the thresholds for different categories of choice observed in our sample. By combining with Tables 3.27 and 3.28, the coefficient in columns (3) is statistically significant at 1% level, which means that the instrumental variables are significant in predicting the endogenous explanatory variables in the first stage. Moreover, for the Sanderson-Windmeijer multivariate F tests of excluded instruments is 104.690 ($p - value = 0.000$) the IV F-stat is 184.20, and Anderson LM statistic is 103.988 ($p - value = 0.000$), which indicate the IVs pass the underidentification test. Cragg-Donald Wald F statistic is equal to 104.692, and according to Stock and Yogo (2005), the 10% maximal size is 16.380, the 104.692 is greater than the critical value, therefore, the joint instrumental variable passes the weak identification test.

Table 3.29: Estimates the Impact of Return Migration Decisions on Life Satisfaction

| VARIABLES | (1) LS | (2) CMP-IV LS | (3) CMP-IV Mig |
|---|---|---|---|
| Distance | | | -0.002*** |
| | | | (0.001) |
| Return_Mig | -0.059* | -0.533*** | |
| | (0.031) | (0.119) | |
| Age | -0.072*** | -0.069*** | -0.175*** |
| | (0.007) | (0.008) | (0.010) |
| Age_sq | 0.088*** | 0.078*** | 0.166*** |
| | (0.008) | (0.009) | (0.011) |
| Edu | -0.003 | -0.008*** | 0.003 |
| | (0.003) | (0.003) | (0.004) |
| Marry | 0.352*** | 0.197*** | 0.027 |
| | (0.038) | (0.035) | (0.049) |
| Male | -0.064** | 0.040* | 0.190*** |
| | (0.026) | (0.023) | (0.032) |
| Lag_LS | -0.009 | 0.323*** | -0.006 |
| | (0.013) | (0.011) | (0.015) |
| ln_Month _Income | -0.049*** | 0.010 | 0.021 |
| | (0.006) | (0.009) | (0.014) |
| ln_AVE_ Social_Network | 0.045*** | 0.031*** | -0.004 |
| | (0.008) | (0.009) | (0.013) |
| ln_CO2 | 0.102*** | -0.015 | -0.112** |
| | (0.039) | (0.037) | (0.053) |
| ln_SO2 | -0.035*** | 0.012 | -0.053** |
| | (0.011) | (0.017) | (0.023) |
| AQI | 0.004*** | -0.004** | 0.003 |
| | (0.001) | (0.002) | (0.004) |

Note: Robust standard errors in parentheses. ***Significant at 1%; **Significant at 5%; *Significant at 10%.

Table 3.30: Continued Table 3.29

| | (1) | (2) | (3) |
|---|---|---|---|
| | | CMP-IV | CMP-IV |
| VARIABLES | LS | LS | Mig |
| PM2_5 | -0.004*** | 0.005** | -0.005 |
| | (0.001) | (0.002) | (0.004) |
| PM10 | -0.000 | 0.001* | 0.001 |
| | (0.000) | (0.001) | (0.001) |
| Greenrate | 0.004* | 0.004 | 0.018*** |
| | (0.002) | (0.003) | (0.005) |
| Constant | | | 1.902*** |
| | | | (0.432) |
| cut_1_1 | -2.855*** | -1.142*** | |
| | (0.251) | (0.328) | |
| cut_1_2 | -1.989*** | -0.551* | |
| | (0.250) | (0.329) | |
| cut_1_3 | -0.814*** | 0.629* | |
| | (0.249) | (0.333) | |
| cut_1_4 | 0.173 | 1.496*** | |
| | (0.249) | (0.337) | |
| N | 30,579 | 30,579 | 30,579 |
| Fixed Effects | YES | YES | YES |

Note: Robust standard errors in parentheses. ***Significant at 1%; **Significant at 5%; *Significant at 10%. The cut points indicate the thresholds for different categories of choice observed in our sample. The constant term is from the first stage of IV estimation. In Tables 3.29 and 3.30, the coefficient is statistically significant at 1% level, which means that the instrumental variables are significant in predicting the endogenous explanatory variables in the first stage. Moreover, for the Sanderson-Windmeijer multivariate F test of excluded instruments is 33.73 ($p-value = 0.000$) the IV F-stat is 104.06, and Anderson LM statistic is 33.688 ($p-value = 0.000$), which indicate the IVs pass the underidentification test. Cragg-Donald Wald F statistic is equal to 33.732, and according to Stock and Yogo (2005), the 10% maximal size is 16.38, the number is greater than the critical value, therefore, the joint instrumental variable passes the weak identification test.

Table 3.31: The Heterogeneous Effects of Education Level on Migration Decisions Through Life Satisfaction

| VARIABLES | (1) CMP-IV Mig | (2) CMP-IV Lag_LS |
|---|---|---|
| Commute_time | | -0.005*** |
| | | (0.001) |
| 1.Educ#c.Commute_time | | 0.003** |
| | | (0.001) |
| 2.Educ#c.Commute_time | | 0.001 |
| | | (0.002) |
| 2.Lag_LS | 0.904*** | |
| | (0.068) | |
| 3.Lag_LS | -0.074 | |
| | (0.067) | |
| 4.Lag_LS | -0.151** | |
| | (0.070) | |
| 5.Lag_LS | -0.189*** | |
| | (0.071) | |
| 1.Educ | -0.144 | -0.038 |
| | (0.097) | (0.032) |
| 2.Educ | 0.294 | 0.168*** |
| | (0.199) | (0.045) |
| 2.Lag_LS#1.Educ | 0.090 | |
| | (0.104) | |
| 2.Lag_LS#2.Educ | -0.511** | |
| | (0.212) | |
| 3.Lag_LS#1.Educ | 0.103 | |
| | (0.102) | |
| 3.Lag_LS#2.Educ | -0.419** | |
| | (0.206) | |
| 4.Lag_LS#1.Educ | 0.061 | |
| | (0.104) | |
| 4.Lag_LS#2.Educ | -0.399* | |
| | (0.206) | |
| 5.Lag_LS#1.Educ | 0.139 | |
| | (0.106) | |
| 5.Lag_LS#2.Educ | -0.449** | |
| | (0.214) | |

Note: Robust standard errors in parentheses. ***Significant at 1%; **Significant at 5%; *Significant at 10%.

Table 3.32: Continued Table 3.11

| VARIABLES | (1) CMP-IV Mig | (2) CMP-IV Lag_LS |
|---|---|---|
| Age | -0.149*** | -0.018*** |
|  | (0.005) | (0.006) |
| Age_sq | 0.132*** | 0.030*** |
|  | (0.006) | (0.007) |
| Marry | 0.012 | 0.209*** |
|  | (0.031) | (0.031) |
| Male | 0.264*** | -0.046*** |
|  | (0.017) | (0.018) |
| Family_Tie | -0.027*** | 0.013*** |
|  | (0.005) | (0.005) |
| ln_Pop_density | -0.062*** | 0.063*** |
|  | (0.011) | (0.011) |
| Unemployrate | -0.071*** | 0.061*** |
|  | (0.026) | (0.022) |
| Industrialratio | 0.051** | -0.058*** |
|  | (0.020) | (0.020) |
| Religion | -0.031 | 0.216** |
|  | (0.079) | (0.085) |
| Hospital | -0.108** | 0.114** |
|  | (0.045) | (0.052) |
| School | 0.074*** | -0.016 |
|  | (0.014) | (0.014) |
| Urban | -0.185*** | 0.113*** |
|  | (0.019) | (0.020) |
| Constant | 2.635*** |  |
|  | (0.148) |  |
| cut_2_1 |  | -1.207*** |
|  |  | (0.148) |
| cut_2_2 |  | -0.259* |
|  |  | (0.147) |
| cut_2_3 |  | 0.729*** |
|  |  | (0.148) |
| cut_2_4 |  | 1.580*** |
|  |  | (0.148) |
| N | 36,680 | 36,680 |
| Fixed Effects | YES | YES |

Note: Robust standard errors in parentheses. ***Significant at 1%; **Significant at 5%; *Significant at 10%. Fixed effects include individual and time effects. The cut points indicate the thresholds for different categories of choice observed in our sample.

Table 3.33: OLS Estimation of Equation 1

| VARIABLES | (1) Mig |
|---|---|
| Lag_LS | -0.007*** |
|  | (0.001) |
| Age | 0.004** |
|  | (0.002) |
| Age_sq | -0.004*** |
|  | (0.001) |
| Edu | 0.002*** |
|  | (0.001) |
| Marry | 0.012*** |
|  | (0.004) |
| Male | -0.003 |
|  | (0.022) |
| Family_Tie | -0.000 |
|  | (0.000) |
| Local_house | -0.009*** |
|  | (0.002) |
| ln_Pop_density | 0.008** |
|  | (0.003) |
| Unemployrate | -0.005*** |
|  | (0.002) |
| Industrialratio | 0.002 |
|  | (0.002) |
| Religion | -0.020*** |
|  | (0.007) |
| Hospital | -0.010*** |
|  | (0.003) |
| School | -0.017*** |
|  | (0.003) |
| Urban | -0.003 |
|  | (0.002) |
| Constant | 0.086 |
|  | (0.073) |
| $N$ | 36,680 |
| Fixed Effects | YES |
| $R^2$ | 0.020 |

Robust standard errors in parentheses. ***Significant at 1%; **Significant at 5%; *Significant at 10%. Fixed effects include individual and time effects.

Table 3.34: OLS Estimation of Equation 2

|  | (1) |
| --- | --- |
| VARIABLES | LS |
| Mig | -0.157** |
|  | (0.069) |
| Age | -0.025** |
|  | (0.012) |
| Age_sq | 0.094*** |
|  | (0.011) |
| Edu | 0.003 |
|  | (0.007) |
| Marry | 0.234*** |
|  | (0.046) |
| Male | 0.276 |
|  | (0.263) |
| Lag_LS | -0.218*** |
|  | (0.007) |
| ln_Month_Income | 0.005 |
|  | (0.005) |
| ln_AVE_Social_Network | 0.008 |
|  | (0.006) |
| ln_CO2 | 0.141 |
|  | (0.086) |
| ln_SO2 | -0.040*** |
|  | (0.007) |
| AQI | 0.000 |
|  | (0.001) |
| PM2_5 | -0.003*** |
|  | (0.001) |
| PM10 | 0.001* |
|  | (0.000) |
| Greenrate | 0.001 |
|  | (0.001) |
| Constant | 3.208*** |
|  | (0.483) |
| $N$ | 30,579 |
| Fixed Effects | YES |
| $R^2$ | 0.117 |

Robust standard errors in parentheses. ***Significant at 1%; **Significant at 5%; *Significant at 10%. Fixed effects include individual and time effects.

Table 3.35: The Full Table: Marginal Effects of the Variables of Interest on Life Satisfaction=1

|  | (1) | |
|---|---|---|
|  | LS | |
|  | b | se |
| Mig | -0.021*** | 0.003 |
| Age | 0.002*** | 0.000 |
| Age_sq | -0.003*** | 0.000 |
| Edu | 0.001*** | 0.000 |
| Marry | -0.016*** | 0.002 |
| Male | 0.002** | 0.001 |
| Lag_LS | -0.022*** | 0.001 |
| ln_Month_Income | -0.001*** | 0.000 |
| ln_AVE_Social_Network | -0.001** | 0.000 |
| ln_CO2 | -0.001 | 0.001 |
| ln_SO2 | -0.001 | 0.000 |
| AQI | -0.000 | 0.000 |
| PM2_5 | 0.000 | 0.000 |
| PM10 | 0.000 | 0.000 |
| Greenrate | 0.000** | 0.000 |
| N | 30,542 | |

Note: Robust standard errors in parentheses. ***Significant at 1%; **Significant at 5%; *Significant at 10%.

Table 3.36: The Full Table: Marginal Effects of the Variables of Interest on Life Satisfaction=2

|  | (1) | |
|---|---|---|
|  | LS | |
|  | b | se |
| Mig | -0.027*** | 0.004 |
| Age | 0.003*** | 0.000 |
| Age_sq | -0.004*** | 0.000 |
| Edu | 0.001*** | 0.000 |
| Marry | -0.021*** | 0.002 |
| Male | 0.003** | 0.001 |
| Lag_LS | -0.029*** | 0.001 |
| ln_Month_Income | -0.002*** | 0.000 |
| ln_AVE_Social_Network | -0.001** | 0.001 |
| ln_CO2 | -0.001 | 0.001 |
| ln_SO2 | -0.001 | 0.001 |
| AQI | -0.000 | 0.000 |
| PM2_5 | 0.000 | 0.000 |
| PM10 | 0.000 | 0.000 |
| Greenrate | 0.000** | 0.000 |
| N | 30,542 | |

Note: Robust standard errors in parentheses. ***Significant at 1%; **Significant at 5%; *Significant at 10%.

Table 3.37: The Full Table: Marginal Effects of the Variables of Interest on Life Satisfaction=3

|  | (1) | |
| --- | --- | --- |
|  | LS | |
|  | b | se |
| Mig | -0.061*** | 0.009 |
| Age | 0.006*** | 0.001 |
| Age_sq | -0.008*** | 0.001 |
| Edu | 0.002*** | 0.000 |
| Marry | -0.048*** | 0.004 |
| Male | 0.006** | 0.003 |
| Lag_LS | -0.067*** | 0.001 |
| ln_Month_Income | -0.003*** | 0.001 |
| ln_AVE_Social_Network | -0.003** | 0.001 |
| ln_CO2 | -0.003 | 0.002 |
| ln_SO2 | -0.002 | 0.001 |
| AQI | -0.000 | 0.000 |
| PM2_5 | 0.000 | 0.000 |
| PM10 | 0.000 | 0.000 |
| Greenrate | 0.000** | 0.000 |
| N | 30,542 | |

Note: Robust standard errors in parentheses. ***Significant at 1%; **Significant at 5%; *Significant at 10%.

Table 3.38: The Full Table: Marginal Effects of the Variables of Interest on Life Satisfaction=4

|  | (1) | |
| --- | --- | --- |
|  | LS | |
|  | b | se |
| Mig | 0.013*** | 0.002 |
| Age | -0.001*** | 0.000 |
| Age_sq | 0.002*** | 0.000 |
| Edu | -0.000*** | 0.000 |
| Marry | 0.010*** | 0.001 |
| Male | -0.001** | 0.001 |
| Lag_LS | 0.014*** | 0.001 |
| ln_Month_Income | 0.001*** | 0.000 |
| ln_AVE_Social_Network | 0.001** | 0.000 |
| ln_CO2 | 0.001 | 0.001 |
| ln_SO2 | 0.000 | 0.000 |
| AQI | 0.000 | 0.000 |
| PM2_5 | -0.000 | 0.000 |
| PM10 | -0.000 | 0.000 |
| Greenrate | -0.000** | 0.000 |
| N | 30,542 | |

Note: Robust standard errors in parentheses. ***Significant at 1%; **Significant at 5%; *Significant at 10%.

Table 3.39: The Full Table: Marginal Effects of the Variables of Interest on Life Satisfaction=5

| | (1) | |
| | LS | |
| | b | se |
|---|---|---|
| Mig | 0.095*** | 0.015 |
| Age | -0.010*** | 0.002 |
| Age_sq | 0.013*** | 0.002 |
| Edu | -0.003*** | 0.000 |
| Marry | 0.075*** | 0.007 |
| Male | -0.010** | 0.004 |
| Lag_LS | 0.104*** | 0.002 |
| ln_Month_Income | 0.005*** | 0.002 |
| ln_AVE_Social_Network | 0.004** | 0.002 |
| ln_CO2 | 0.004 | 0.004 |
| ln_SO2 | 0.003 | 0.002 |
| AQI | 0.000 | 0.000 |
| PM2_5 | -0.000 | 0.000 |
| PM10 | -0.000 | 0.000 |
| Greenrate | -0.000** | 0.000 |
| N | 30,542 | |

Note: Robust standard errors in parentheses. ***Significant at 1%; **Significant at 5%; *Significant at 10%.

Table 3.40: The Full Table: Marginal Effects of the Return Migration Decision on Life Satisfaction=1

| | (1) | |
| | LS | |
| | b | se |
|---|---|---|
| Return_Mig | 0.045*** | 0.011 |
| Age | 0.006*** | 0.001 |
| Age_sq | -0.007*** | 0.001 |
| Edu | 0.001*** | 0.000 |
| Marry | -0.017*** | 0.003 |
| Male | -0.003* | 0.002 |
| Lag_LS | -0.027*** | 0.001 |
| ln_Month_Income | -0.001 | 0.001 |
| ln_AVE_Social_Network | -0.003*** | 0.001 |
| ln_CO2 | 0.001 | 0.003 |
| ln_SO2 | -0.001 | 0.001 |
| AQI | 0.000** | 0.000 |
| PM2_5 | -0.000** | 0.000 |
| PM10 | -0.000* | 0.000 |
| Greenrate | -0.000 | 0.000 |
| N | 15,471 | |

Note: Robust standard errors in parentheses. ***Significant at 1%; **Significant at 5%; *Significant at 10%.

Table 3.41: The Full Table: Marginal Effects of the Return Migration Decision on Life Satisfaction=2

|  | (1) | |
| --- | --- | --- |
|  | LS | |
|  | b | se |
| Return_Mig | 0.049*** | 0.011 |
| Age | 0.006*** | 0.001 |
| Age_sq | -0.007*** | 0.001 |
| Edu | 0.001*** | 0.000 |
| Marry | -0.018*** | 0.003 |
| Male | -0.004* | 0.002 |
| Lag_LS | -0.030*** | 0.001 |
| ln_Month_Income | -0.001 | 0.001 |
| ln_AVE_Social_Network | -0.003*** | 0.001 |
| ln_CO2 | 0.001 | 0.003 |
| ln_SO2 | -0.001 | 0.002 |
| AQI | 0.000** | 0.000 |
| PM2_5 | -0.000** | 0.000 |
| PM10 | -0.000* | 0.000 |
| Greenrate | -0.000 | 0.000 |
| N | 15,471 | |

Note: Robust standard errors in parentheses. ***Significant at 1%; **Significant at 5%; *Significant at 10%.

Table 3.42: The Full Table: Marginal Effects of the Return Migration Decision on Life Satisfaction=3

|  | (1) | |
| --- | --- | --- |
|  | LS | |
|  | b | se |
| Return_Mig | 0.092*** | 0.018 |
| Age | 0.012*** | 0.001 |
| Age_sq | -0.013*** | 0.001 |
| Edu | 0.001*** | 0.000 |
| Marry | -0.034*** | 0.006 |
| Male | -0.007* | 0.004 |
| Lag_LS | -0.056*** | 0.003 |
| ln_Month_Income | -0.002 | 0.002 |
| ln_AVE_Social_Network | -0.005*** | 0.002 |
| ln_CO2 | 0.003 | 0.006 |
| ln_SO2 | -0.002 | 0.003 |
| AQI | 0.001** | 0.000 |
| PM2_5 | -0.001** | 0.000 |
| PM10 | -0.000* | 0.000 |
| Greenrate | -0.001 | 0.001 |
| N | 15,471 | |

Note: Robust standard errors in parentheses. ***Significant at 1%; **Significant at 5%; *Significant at 10%.

Table 3.43: The Full Table: Marginal Effects of the Return Migration Decision on Life Satisfaction=4

|  | (1) LS | |
| --- | --- | --- |
|  | b | se |
| Return_Mig | -0.032*** | 0.006 |
| Age | -0.004*** | 0.000 |
| Age_sq | 0.005*** | 0.001 |
| Edu | -0.000*** | 0.000 |
| Marry | 0.012*** | 0.002 |
| Male | 0.002* | 0.001 |
| Lag_LS | 0.019*** | 0.001 |
| ln_Month_Income | 0.001 | 0.001 |
| ln_AVE_Social_Network | 0.002*** | 0.001 |
| ln_CO2 | -0.001 | 0.002 |
| ln_SO2 | 0.001 | 0.001 |
| AQI | -0.000** | 0.000 |
| PM2_5 | 0.000** | 0.000 |
| PM10 | 0.000* | 0.000 |
| Greenrate | 0.000 | 0.000 |
| N | 15,471 | |

Note: Robust standard errors in parentheses. ***Significant at 1%; **Significant at 5%; *Significant at 10%.

Table 3.44: The Full Table: Marginal Effects of the Return Migration Decision on Life Satisfaction=5

|  | (1) LS | |
| --- | --- | --- |
|  | b | se |
| Return_Mig | -0.154*** | 0.034 |
| Age | -0.020*** | 0.002 |
| Age_sq | 0.023*** | 0.003 |
| Edu | -0.002*** | 0.001 |
| Marry | 0.057*** | 0.010 |
| Male | 0.011* | 0.007 |
| Lag_LS | 0.094*** | 0.003 |
| ln_Month_Income | 0.003 | 0.003 |
| ln_AVE_Social_Network | 0.009*** | 0.003 |
| ln_CO2 | -0.004 | 0.011 |
| ln_SO2 | 0.004 | 0.005 |
| AQI | -0.001** | 0.001 |
| PM2_5 | 0.001** | 0.001 |
| PM10 | 0.000* | 0.000 |
| Greenrate | 0.001 | 0.001 |
| N | 15,471 | |

Note: Robust standard errors in parentheses. ***Significant at 1%; **Significant at 5%; *Significant at 10%.

Table 3.45: IV-OLS Estimation of Equation 1

| VARIABLES | (1)<br>First<br>Lag_LS | (2)<br>Second<br>Mig | (3)<br>First<br>Lag_LS | (4)<br>Second<br>Mig | (5)<br>First<br>Lag_LS | (6)<br>Second<br>Mig |
|---|---|---|---|---|---|---|
| Overtime_work | -0.002*** | | | | -0.002** | |
| | (0.000) | | | | (0.001) | |
| Commute_time_ | | | -0.003*** | | -0.003*** | |
| | | | (0.001) | | (0.001) | |
| Lag_LS | | -0.184** | | -0.307*** | | -0.282*** |
| | | (0.085) | | (0.093) | | (0.078) |
| Age | -0.011*** | -0.049*** | -0.016*** | -0.056*** | -0.015** | -0.056*** |
| | (0.004) | (0.002) | (0.006) | (0.003) | (0.006) | (0.003) |
| Age_sq | 0.019*** | 0.046*** | 0.026*** | 0.054*** | 0.025*** | 0.054*** |
| | (0.004) | (0.002) | (0.007) | (0.004) | (0.007) | (0.004) |
| Edu | 0.003** | -0.002*** | 0.011*** | -0.001 | 0.010*** | -0.002 |
| | (0.001) | (0.001) | (0.002) | (0.001) | (0.002) | (0.001) |
| Marry | 0.230*** | 0.026 | 0.202*** | 0.073*** | 0.200*** | 0.069*** |
| | (0.022) | (0.021) | (0.031) | (0.023) | (0.031) | (0.021) |
| Male | -0.074*** | 0.055*** | -0.051*** | 0.058*** | -0.044** | 0.061*** |
| | (0.012) | (0.008) | (0.017) | (0.009) | (0.017) | (0.009) |
| Family_Tie | 0.007* | -0.007*** | 0.012** | -0.007*** | 0.009* | -0.007*** |
| | (0.004) | (0.002) | (0.005) | (0.002) | (0.005) | (0.002) |
| ln_Pop_density | 0.045*** | -0.011** | 0.061*** | -0.022*** | 0.058*** | -0.024*** |
| | (0.008) | (0.005) | (0.011) | (0.007) | (0.011) | (0.007) |
| Unemployrate | 0.044*** | -0.010 | 0.049** | -0.011 | 0.048** | -0.012 |
| | (0.017) | (0.007) | (0.021) | (0.011) | (0.021) | (0.010) |
| Industrialratio | -0.037*** | 0.005 | -0.023 | -0.005 | -0.028 | -0.004 |
| | (0.013) | (0.006) | (0.020) | (0.009) | (0.020) | (0.009) |
| Religion | 0.162*** | 0.024 | 0.250*** | 0.040 | 0.249*** | 0.033 |
| | (0.054) | (0.024) | (0.080) | (0.043) | (0.080) | (0.040) |
| Hospital | 0.100*** | -0.021 | 0.061 | -0.033 | 0.063 | -0.033 |
| | (0.031) | (0.014) | (0.052) | (0.023) | (0.052) | (0.023) |
| School | 0.005 | 0.025*** | -0.008 | 0.035*** | -0.009 | 0.036*** |
| | (0.010) | (0.003) | (0.014) | (0.006) | (0.014) | (0.006) |
| Urban | 0.107*** | -0.029*** | 0.105*** | 0.010 | 0.114*** | 0.008 |
| | (0.013) | (0.010) | (0.020) | (0.013) | (0.020) | (0.012) |
| Constant | 2.653*** | 1.941*** | 2.786*** | 2.505*** | 2.853*** | 2.434*** |
| | (0.098) | (0.227) | (0.141) | (0.260) | (0.143) | (0.222) |

Robust standard errors in parentheses. ***Significant at 1%; **Significant at 5%; *Significant at 10%.

Table 3.46: Continued Table 3.45

| VARIABLES | (1) First Lag_LS | (2) Second Mig | (3) First Lag_LS | (4) Second Mig | (5) First Lag_LS | (6) Second Mig |
|---|---|---|---|---|---|---|
| $N$ | 36,680 | 36,680 | 36,680 | 36,680 | 36,680 | 36,680 |
| Fixed Effects | YES | YES | YES | YES | YES | YES |
| $R^2$ | 0.063 | 0.017 | 0.081 | 0.208 | 0.082 | 0.133 |
| IV F-stat | | 235.22 | | 105.10 | | 110.51 |
| Sanderson-Windmeijer F test statistic | | 14.548 | | 22.740 | | 15.348 |
| Anderson LM | | 14.548 | | 22.732 | | 30.676 |
| Cragg-Donald Wald F statistic | | 14.545 | | 22.736 | | 15.348 |
| Hansen Statistic | | N/A | | N/A | | 0.069 |
| Hansen p-value | | N/A | | N/A | | 0.793 |

Robust standard errors in parentheses. ***Significant at 1%; **Significant at 5%; *Significant at 10%. Fixed effects include individual and time effects.

In Tables 3.47 and 3.48, I report the IV specifications. Column 1 is the first stage of 2SLS results, and column 2 is the second stage results. The coefficients in the first stage are statistically significant at 1% level, which means that the instrumental variable is significant in predicting the endogenous explanatory variables in the first stage. For the Sanderson-Windmeijer multivariate F test of excluded instruments is 83.250, all with respect to $p - value = 0.000$, which rejects the null that the particular endogenous regressor in question is unidentified. The IV F-stat is 226.880, and Anderson LM statistic is 82.910 with respect to $p - value = 0.000$, which indicates the IV passes the underidentification test. Cragg-Donald Wald F statistic is equal to 83.09, according to Hausman et al. (2005); Stock and Yogo (2005), the 10% maximal size of the Stock-Yogo weak ID test critical value is 16.38. This indicates IV does not fall to weak identification.

Table 3.47: IV-OLS Estimation of Equation 2

| VARIABLES | (1) First Mig | (2) Second LS |
|---|---|---|
| Distance | -0.001*** | |
| | (0.000) | |
| Mig | -0.643** | |
| | (0.319) | |
| Age | -0.050*** | -0.073*** |
| | (0.001) | (0.016) |
| Age_sq | 0.045*** | 0.078*** |
| | (0.001) | (0.015) |
| Edu | -0.000 | -0.007*** |
| | (0.000) | (0.001) |
| Marry | -0.028*** | 0.214*** |
| | (0.007) | (0.023) |
| Male | 0.066*** | 0.032 |
| | (0.004) | (0.024) |
| Lag_LS | -0.022*** | 0.260*** |
| | (0.002) | (0.009) |
| ln_Month_Income | 0.000 | 0.009* |
| | (0.002) | (0.005) |
| ln_AVE_Social_Network | -0.007*** | 0.008 |
| | (0.002) | (0.006) |
| ln_CO2 | -0.020*** | 0.035*** |
| | (0.004) | (0.013) |
| ln_SO2 | 0.013*** | -0.020*** |
| | (0.002) | (0.007) |
| AQI | -0.000 | 0.001 |
| | (0.000) | (0.001) |
| PM2_5 | -0.000 | -0.004*** |
| | (0.000) | (0.001) |
| PM10 | 0.000 | 0.001*** |
| | (0.000) | (0.000) |
| Greenrate | -0.000 | -0.001 |
| | (0.000) | (0.001) |
| Constant | 1.231*** | 4.147*** |
| | (0.041) | (0.410) |

Robust standard errors in parentheses. ***Significant at 1%; **Significant at 5%; *Significant at 10%.

Table 3.48: Continued Table 3.47

| VARIABLES | (1) First Mig | (2) Second LS |
|---|---|---|
| $N$ | 30,579 | 30,579 |
| Fixed Effects | YES | YES |
| $R^2$ | 0.184 | 0.074 |
| IV F-stat | | 226.88 |
| Sanderson-Windmeijer F test statistic | | 83.250 |
| Anderson LM | | 82.91 |
| Cragg-Donald Wald F statistic | | 83.09 |

Robust standard errors in parentheses. ***Significant at 1%; **Significant at 5%; *Significant at 10%. Fixed effects include individual and time effects.

## 3.A.2 Additional Equations

### 3.A.2.1 Mathematical Approach to Proof Endogenous Issue

In a simultaneous equation system situation, the structural equations are set, as Equation (3.1) and Equation (3.2), as follows:

$$Freq\_mig_{i,t} = \alpha_0 + \alpha_1 \Delta LS_{i,t} + \boldsymbol{\alpha_2 Exo\_Var_{1i,t}} + \mu_i + \gamma_t + \epsilon_{i,t} \qquad (3.7)$$

$$\Delta LS_{i,t} = \beta_0 + \beta_1 Freq\_mig_{i,t} + \boldsymbol{\beta_2 Exo\_Var_{2i,t}} + \eta_i + \lambda_t + \nu_{i,t} \qquad (3.8)$$

where the vectors of $\boldsymbol{Exo\_Var_{1i,t}}$ and $\boldsymbol{Exo\_Var_{2i,t}}$ include the control variables, which are exogenous, from demographic ($\boldsymbol{X_{1i,t}}$, $\boldsymbol{X_{2i,t}}$), and city level ($\boldsymbol{City_{1i,t}}$, $\boldsymbol{City_{2i,t}}$), respectively. $\boldsymbol{\alpha_2}$ and $\boldsymbol{\beta_2}$ are the vectors of corresponding coefficients. $\mu_i$ and $\eta_i$ capture the individual effect; $\gamma_t$ and $\lambda_t$ capture the time effect; and $\epsilon_{i,t}$ and $\nu_{i,t}$ are the error terms. $\alpha_0$ and $\beta_0$ are constant terms.

$$E(\epsilon_i, t | \boldsymbol{Exo\_Var_{1i,t} Exo\_Var_{2i,t}}) = 0 \qquad (3.9)$$

$$E(\nu_i, t | \boldsymbol{Exo\_Var_{1i,t} Exo\_Var_{2i,t}}) = 0 \qquad (3.10)$$

If $\epsilon_{i,t}$ and $\nu_{i,t}$ are independent of each other and assuming $\alpha_1 \beta_1 \neq 1$, this means that two structural equations do not describe two endogenous variables having the same structural relationship. Then we can have the following two corollaries:

Corollary 1) if $\alpha \neq 0$, $E(\nu_{i,t} | Freq\_mig_{i,t} \neq 0)$ or constant.

Corollary 2) if $\beta_1 \neq 0$, $E(\epsilon_{i,t} | \Delta LS_{i,t} \neq 0)$ or constant.

Taking Equation (3.7) into Equation (3.8),

$$\Delta LS_{i,t} = \beta_0 + \beta_1(\alpha_0 + \alpha_1 \Delta LS_{i,t} + \boldsymbol{\alpha_2 Exo\_Var_{1i,t}} + \mu_i + \gamma_t + \epsilon_{i,t})$$
$$+ \boldsymbol{\beta_2 Exo\_Var_{2i,t}} + \eta_i + \lambda_t + \nu_{i,t} \tag{3.11}$$

Reduced form:

$$\Delta LS_{i,t} = \frac{\beta_0 + \beta_1\alpha_0}{1 - \beta_1\alpha_1} + \frac{\beta_1\alpha_2}{1 - \beta_1\alpha_1}\boldsymbol{Exo\_Var_{1i,t}} + \frac{\beta_2}{1 - \beta_1\alpha_1}\boldsymbol{Exo\_Var_{2i,t}}$$
$$+ \frac{\beta_1}{1 - \beta_1\alpha_1}\epsilon_{i,t} + \frac{\nu_{i,t}}{1 - \beta_1\alpha_1} \tag{3.12}$$

$$E\left(\Delta LS_{i,t}\epsilon_{i,t}\right) = \frac{E(\beta_1\epsilon_{i,t}^2)}{1 - \beta_1\alpha_1} = \frac{\beta_1\sigma_{\epsilon_{i,t}}^2}{1 - \beta_1\alpha_1} \neq 0$$

In a similar way, solving $Freq\_mig_{i,t}$, taking Equation (3.8) into Equation (3.7),

$$Freq\_mig_{i,t} = \alpha_0 + \alpha_1(\beta_0 + \beta_1 Freq\_mig_{i,t} + \boldsymbol{\beta_2 Exo\_Var_{2i,t}} + \eta_i + \lambda_t + \nu_{i,t})$$
$$+ \boldsymbol{\alpha_2 Exo\_Var_{1i,t}} + \mu_i + \gamma_t + \epsilon_{i,t} \tag{3.13}$$

Reduced form:

$$Freq\_mig_{i,t} = \frac{\alpha_0 + \alpha_1\beta_1}{1 - \beta_1\alpha_1} + \frac{\beta_2\alpha_1}{1 - \beta_1\alpha_1}\boldsymbol{Exo\_Var_{1i,t}} + \frac{\alpha_2}{1 - \beta_1\alpha_1}\boldsymbol{Exo\_Var_{2i,t}}$$
$$+ \frac{\alpha_1\nu_{i,t}}{1 - \beta_1\alpha_1} + \frac{\epsilon_{i,t}}{1 - \beta_1\alpha_1} \tag{3.14}$$

$$E(Freq\_mig_{i,t}\nu_{i,t}) = \frac{E(\alpha_1\nu_{i,t}^2)}{1 - \beta_1\alpha_1} = \frac{\alpha_1\sigma_{\nu_{i,t}}^2}{1 - \beta_1\alpha_1} \neq 0$$

Again, given the above proof of the bidirectional causation, if $\epsilon_{i,t}$ is positive, then it causes $Freq\_mig_{i,t}$ bigger. However, this bidirectional causation tells us that bigger $Freq\_mig_{i,t}$ would influence $\Delta LS_{i,t}$. So, if $\beta_1$ is positive or negative and $Freq\_mig_{i,t}$ is bigger, $\Delta LS_{i,t}$ and $\Delta LS_{i,t}$ and $\nu_{i,t}$ will have a positive or negative correlation. We can have the following covariance between $\Delta LS_{i,t}$ and $\epsilon_{i,t}$:

$$Cov(\Delta LS_{i,t}, \epsilon_{i,t}) = Cov(\beta_0 + \beta_1 Freq\_mig_{i,t} + \boldsymbol{\beta_2 Exo\_Var_{2i,t}} + \nu_{i,t}, \epsilon_{i,t})$$
$$= \beta_1 Cov(Freq\_mig_{i,t}, \epsilon_{i,t}) + Cov(\nu_{i,t}, \epsilon_{i,t}) = \beta_1 \frac{\sigma_{\epsilon_{i,t}}^2}{1 - \beta_1\alpha_1} \tag{3.15}$$

Therefore, from the above proving process, we can understand that simultaneous endogeneity exists between the change in life satisfaction and the frequency of migration mobility.

# Bibliography

Acemoglu, D. and Angrist, J. (2000), 'How large are human-capital externalities? evidence from compulsory schooling laws', *NBER macroeconomics annual* **15**, 9–59.

Ahmad, A. N. (2016), *Masculinity, sexuality and illegal migration: Human smuggling from Pakistan to Europe*, 1st edn, Routledge.

Altonji, J. G. and Card, D. (1991), The effects of immigration on the labor market outcomes of less-skilled natives, *in* 'Immigration, trade, and the labor market', University of Chicago Press, pp. 201–234.

Andreenkova, A. and Andreenkova, N. (2019), 'Comparative analysis of life satisfaction determinants in russia and in europe', *Discourse* **5**(1), 67–81.

Arabmazar, A. and Schmidt, P. (1981), 'Further evidence on the robustness of the tobit estimator to heteroskedasticity', *Journal of Econometrics* **17**(2), 253–258.

Aristei, D., Perali, F. and Pieroni, L. (2008), 'Cohort, age and time effects in alcohol consumption by italian households: a double-hurdle approach', *Empirical Economics* **35**, 29–61.

Aristei, D. and Pieroni, L. (2008), 'A double-hurdle approach to modelling tobacco consumption in italy', *Applied Economics* **40**(19), 2463–2476.

Arrondel, L. and Savignac, F. (2015), 'Risk management, housing and stockholding', *Applied Economics* **47**(39), 4208–4227.

Au, C.-C. and Henderson, J. V. (2006), 'Are chinese cities too small?', *The Review of Economic Studies* **73**(3), 549–576.

Autor, D. H. and 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.

Aydemir, A. and Borjas, G. J. (2006), 'A comparative analysis of the labor market impact of international migration: Canada, mexico, and the united states'.

Bacon, P. M., Conte, A. and Moffatt, P. G. (2020), 'A test of risk vulnerability in the wider population', *Theory and Decision* **88**(1), 37–50.

Baek, E. and DeVaney, S. A. (2005), 'Human capital, bequest motives, risk, and the purchase of life insurance', *Journal of Personal Finance* **4**(2), 62–84.

Baptista, A. M. (2008), 'Optimal delegated portfolio management with background risk', *Journal of Banking & Finance* **32**(6), 977–985.

Bartolini, S. (2007), 'Why are people so unhappy? why do they strive so hard for money? competing explanations of the broken promises of economic growth', *Handbook on the Economics of Happiness* pp. 337–64.

Bartolini, S. (2011), 'Manifesto for happiness. shifting society from money to well-being'.

Bartram, D. (2011), 'Economic migration and happiness: Comparing immigrants' and natives' happiness gains from income', *Social Indicators Research* **103**, 57–76.

Bartram, D. (2013), 'Happiness and 'economic migration': A comparison of eastern european migrants and stayers', *Migration studies* **1**(2), 156–175.

Bartram, D. (2015), 'Migration and quality of life in the global context', *Global handbook of quality of life: Exploration of well-being of nations and continents* pp. 491–503.

Baykara-Krumme, H. and Platt, L. (2018), 'Life satisfaction of migrants, stayers and returnees: Reaping the fruits of migration in old age?', *Ageing & Society* **38**(4), 721–745.

Bertaut, C. C. (1998), 'Stockholding behavior of us households: Evidence from the 1983–1989 survey of consumer finances', *Review of Economics and Statistics* **80**(2), 263–275.

Betermier, S., Jansson, T., Parlour, C. and Walden, J. (2012), 'Hedging labor income risk', *Journal of Financial Economics* **105**(3), 622–639.

Bijwaard, G. E. (2015), 'Income of immigrants and their return', *IZA World of Labor* .

Bijwaard, G. E. and Wahba, J. (2014), 'Do high-income or low-income immigrants leave faster?', *Journal of Development Economics* **108**, 54–68.

Blanchflower, D. G. (2009), International evidence on well-being, *in* 'Measuring the subjective well-being of nations: National accounts of time use and well-being', University of Chicago Press, pp. 155–226.

Blanchflower, D. G. and Oswald, A. J. (2004), 'Well-being over time in britain and the usa', *Journal of public economics* **88**(7-8), 1359–1386.

Blau, F. D. and Kahn, L. M. (2015), Immigration and the distribution of incomes, *in* 'Handbook of the economics of international migration', Vol. 1, Elsevier, pp. 793–843.

Borjas, G. J. (2003), 'The labor demand curve is downward sloping: Reexamining the impact of immigration on the labor market', *The Quarterly Journal of Economics* **118**(4), 1335–1374.

Borjas, G. J., Freeman, R. B., Katz, L. F., DiNardo, J. and Abowd, J. M. (1997), 'How much do immigration and trade affect labor market outcomes?', *Brookings Papers on Economic Activity* **1997**(1), 1–90.

Borjas, G. J. and Van Ours, J. C. (2010), *Labor economics*, McGraw-Hill/Irwin Boston.

Brueckner, J. K. (1997), 'Consumption and investment motives and the portfolio choices of homeowners', *The Journal of Real Estate Finance and Economics* **15**, 159–180.

Burstein, A., Hanson, G., Tian, L. and Vogel, J. (2020), 'Tradability and the labor-market impact of immigration: Theory and evidence from the united states', *Econometrica* **88**(3), 1071–1112.

Burton, M., Tomlinson, M. and Young, T. (1994), 'Consumers 'decisions whether or not to purchase meat: A double hurdle analysis of single adult households', *Journal of Agricultural Economics* **45**(2), 202–212.

Cai, M., Shi, J., Ni, Y. and Pan, R. (2013), 'Do heterogeneous background risks matter to household asset allocation?', *Asia-Pacific Journal of Financial Studies* **42**(4), 563–589.

Cai, R., Esipova, N., Oppenheimer, M. and Feng, S. (2014), 'International migration desires related to subjective well-being', *IZA Journal of Migration* **3**, 1–20.

Campbell, J. Y. (2006), 'Household finance', *The Journal of Finance* **61**(4), 1553–1604.

Campbell, J. Y., Chan, Y. L. and Viceira, L. M. (2003), 'A multivariate model of strategic asset allocation', *Journal of Financial Economics* **67**(1), 41–80.

Bibliography

Campbell, J. Y. and Viceira, L. M. (2002), *Strategic asset allocation: portfolio choice for long-term investors*, Clarendon Lectures in Economic.

Card, D. (1990), 'The impact of the mariel boatlift on the miami labor market', *ILR Review* **43**(2), 245–257.

Card, D. (2001), 'Immigrant inflows, native outflows, and the local labor market impacts of higher immigration', *Journal of Labor Economics* **19**(1), 22–64.

Card, D. (2009), 'Immigration and inequality', *American Economic Review* **99**(2), 1–21.

Card, D. and Lewis, E. G. (2007), The diffusion of mexican immigrants during the 1990s: Explanations and impacts, *in* 'Mexican Immigration to the United States', University of Chicago Press, pp. 193–228.

Cardak, B. A. and Wilkins, R. (2009), 'The determinants of household risky asset holdings: Australian evidence on background risk and other factors', *Journal of banking & Finance* **33**(5), 850–860.

Chang, C.-c., Oh, R. R. Y., Le Nghiem, T. P., Zhang, Y., Tan, C. L., Lin, B. B., Gaston, K. J., Fuller, R. A. and Carrasco, L. R. (2020), 'Life satisfaction linked to the diversity of nature experiences and nature views from the window', *Landscape and Urban Planning* **202**, 103874.

Charness, G. and Gneezy, U. (2010), 'Portfolio choice and risk attitudes: An experiment', *Economic Inquiry* **48**(1), 133–146.

Chauvin, J. P., Glaeser, E., Ma, Y. and Tobio, K. (2017), 'What is different about urbanization in rich and poor countries? cities in brazil, china, india and the united states', *Journal of Urban Economics* **98**, 17–49.

Chen, M., Gong, Y., Li, Y., Lu, D. and Zhang, H. (2016), 'Population distribution and urbanization on both sides of the hu huanyong line: Answering the premier's question', *Journal of Geographical Sciences* **26**, 1593–1610.

Chen, Y. and Rosenthal, S. S. (2008), 'Local amenities and life-cycle migration: Do people move for jobs or fun?', *Journal of Urban Economics* **64**(3), 519–537.

Cheng, C., Cheung, M. W.-L., Montasem, A. and of Well-Being Studies, I. N. (2016), 'Explaining differences in subjective well-being across 33 nations using multilevel models: Universal personality, cultural relativity, and national income', *Journal of personality* **84**(1), 46–58.

Chetty, R., Sándor, L. and Szeidl, A. (2017), 'The effect of housing on portfolio choice', *The Journal of Finance* **72**(3), 1171–1212.

Chetty, R. and Szeidl, A. (2007), 'Consumption commitments and risk preferences', *The Quarterly Journal of Economics* **122**(2), 831–877.

Cheung, F. and Lucas, R. E. (2014), 'Assessing the validity of single-item life satisfaction measures: Results from three large samples', *Quality of Life research* **23**, 2809–2818.

Chindarkar, N. (2014), 'Is subjective well-being of concern to potential migrants from latin america?', *Social Indicators Research* **115**, 159–182.

Chiswick, C. U., Chiswick, B. R. and Karras, G. (1992), The impact of immigrants on the macroeconomy, *in* 'Carnegie-Rochester Conference Series on Public Policy', Vol. 37, Elsevier, pp. 279–316.

Chuankai, Y. and Yuemin, N. (2015), 'Evolution of spatial pattern of inter-provincial migration and its impacts on urbanization in china', *Geographical Research* **34**(8), 1492–1506.

Cocco, J. F. (2005), 'Portfolio choice in the presence of housing', *The Review of Financial Studies* **18**(2), 535–567.

Cocco, J. F., Gomes, F. J. and Maenhout, P. J. (2005), 'Consumption and portfolio choice over the life cycle', *The Review of Financial Studies* **18**(2), 491–533.

Cohen-Goldner, S. and Paserman, M. D. (2011), 'The dynamic impact of immigration on natives' labor market outcomes: Evidence from israel', *European Economic Review* **55**(8), 1027–1045.

Combes, P.-P., Démurger, S. and Li, S. (2013), 'Urbanisation and migration externalities in china', *SSRN* .

Combes, P.-P., Démurger, S. and Li, S. (2015), 'Migration externalities in chinese cities', *European Economic Review* **76**, 152–167.

Combes, P.-P., Démurger, S., Li, S. and Wang, J. (2020), 'Unequal migration and urbanisation gains in china', *Journal of Development Economics* **142**, 102328.

Constant, A. and Massey, D. S. (2002), 'Return migration by german guestworkers: Neoclassical versus new economic theories', *International migration* **40**(4), 5–38.

Cortes, P. and Tessada, J. (2011), 'Low-skilled immigration and the labor supply of highly skilled women', *American Economic Journal: Applied Economics* **3**(3), 88–123.

Cragg, J. G. (1971), 'Some statistical models for limited dependent variables with application to the demand for durable goods', *Econometrica* **39**(5), 829–844.

Curcuru, S., Heaton, J., Lucas, D. and Moore, D. (2010), Heterogeneity and portfolio choice: Theory and evidence, *in* 'Handbook of financial econometrics: Tools and techniques', Elsevier, pp. 337–382.

Davin, D. (1998), *Internal migration in contemporary China*, Springer.

Davis, S. J. and Willen, P. (2000), Using financial assets to hedge labor income risks: estimating the benefits, Technical report, Citeseer.

De Jong, G. F., Chamratrithirong, A. and Tran, Q.-G. (2002), 'For better, for worse: Life satisfaction consequences of migration 1', *International Migration Review* **36**(3), 838–863.

Decker, S. and Schmitz, H. (2015), '801 soeppapers-health shocks and risk aversion'.

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

Di Maria, C., Lange, I. and Lazarova, E. (2018), 'A look upstream: Market restructuring, risk, procurement contracts and efficiency', *International Journal of Industrial Organization* **57**, 35–83.

Diener, E., Emmons, R. A., Larsen, R. J. and Griffin, S. (1985), 'The satisfaction with life scale', *Journal of Personality Assessment* **49**(1), 71–75.

Diener, E. and Fujita, F. (1995), 'Resources, personal strivings, and subjective well-being: a nomothetic and idiographic approach.', *Journal of personality and social psychology* **68**(5), 926.

Diener, E., Inglehart, R. and Tay, L. (2013), 'Theory and validity of life satisfaction scales', *Social indicators research* **112**, 497–527.

Diener, E., Suh, E. M., Lucas, R. E. and Smith, H. L. (1999), 'Subjective well-being: Three decades of progress.', *Psychological bulletin* **125**(2), 276.

Dionne, G., Artís, M. and Guillén, M. (1996), 'Count data models for a credit scoring system', *Journal of Empirical Finance* **3**(3), 303–325.

Dolan, P., Peasgood, T. and White, M. (2008), 'Do we really know what makes us happy? a review of the economic literature on the factors associated with subjective well-being', *Journal of economic psychology* **29**(1), 94–122.

Dröes, M. I. and Hassink, W. H. (2013), 'House price risk and the hedging benefits of home ownership', *Journal of Housing Economics* **22**(2), 92–99.

Du, Y., Park, A. and Wang, S. (2005), 'Migration and rural poverty in china', *Journal of comparative economics* **33**(4), 688–709.

Dustmann, C. (1996), 'Return migration: the european experience', *Economic policy* **11**(22), 213–250.

Dustmann, C., Frattini, T. and Preston, I. P. (2013), 'The effect of immigration along the distribution of wages', *Review of Economic Studies* **80**(1), 145–173.

Dynan, K. E., Skinner, J. and Zeldes, S. P. (2004), 'Do the rich save more?', *Journal of Political Economy* **112**(2), 397–444.

Eakins, J. (2016), 'An application of the double hurdle model to petrol and diesel household expenditures in ireland', *Transport Policy* **47**, 84–93.

Easterlin, R. A. (2001), 'Income and happiness: Towards a unified theory', *The economic journal* **111**(473), 465–484.

Easterlin, R. A. and Angelescu, L. (2009), 'Happiness and growth the world over: Time series evidence on the happiness-income paradox', *SSRN* .

Edwards, R. D. (2008), 'Health risk and portfolio choice', *Journal of Business & Economic Statistics* **26**(4), 472–485.

Elmendorf, D. W. and Kimball, M. S. (2000), 'Taxation of labor income and the demand for risky assets', *International Economic Review* **41**(3), 801–832.

Engel, C. and Moffatt, P. G. (2014), 'dhreg, xtdhreg, and bootdhreg: Commands to implement double-hurdle regression', *The Stata Journal* **14**(4), 778–797.

Erlinghagen, M. (2016), Anticipation of life satisfaction before emigration: Evidence from german panel data, *in* 'Advances in Happiness Research: A Comparative Perspective', Springer, pp. 229–244.

Ettema, D., Gärling, T., Olsson, L. E. and Friman, M. (2010), 'Out-of-home activities, daily travel, and subjective well-being', *Transportation Research Part A: Policy and Practice* **44**(9), 723–732.

Fan, C. C. (2011), 'Settlement intention and split households: Findings from a survey of migrants in beijing's urban villages', *China Review* **11**(2), 11–41.

Faßhauer, A. and Rehdanz, K. (2015), Estimating benefits from regional amenities: Internal migration and life satisfaction, Technical report, SOEPpapers on Multidisciplinary Panel Data Research.

Findlay, A. M. and Nowok, B. (2012), 'The uneven impact of different life domains on the wellbeing of migrants'.

Franke, G., Schlesinger, H. and Stapleton, R. C. (2006), 'Multiplicative background risk', *Management Science* **52**(1), 146–153.

Fratantoni, M. C. (1998), 'Homeownership and investment in risky assets', *Journal of Urban Economics* **44**(1), 27–42.

Fratantoni, M. C. (2001), 'Homeownership, committed expenditure risk, and the stockholding puzzle', *Oxford Economic Papers* **53**(2), 241–259.

Frey, B. S. (2018), *Economics of happiness*, Springer.

Friedberg, R. M. (2001), 'The impact of mass migration on the israeli labor market', *The Quarterly Journal of Economics* **116**(4), 1373–1408.

Fuchs-Schündeln, N. and Schündeln, M. (2009), 'Who stays, who goes, who returns? east–west migration within germany since reunification 1', *Economics of Transition* **17**(4), 703–738.

Ghahreman, Y. (2016), *Essays in Entrepreneurship and Household Finance*, Washington University in St. Louis.

Glatz, C. and Eder, A. (2020), 'Patterns of trust and subjective well-being across europe: New insights from repeated cross-sectional analyses based on the european social survey 2002–2016', *Social Indicators Research* **148**(2), 417–439.

Goetzmann, W. N. and Valaitis, E. (2006), 'Simulating real estate in the investment portfolio: model uncertainty and inflation hedging', *Social Science Research Network* .

Golden, L. and Wiens-Tuers, B. (2006), 'To your happiness? extra hours of labor supply and worker well-being', *The Journal of Socio-Economics* **35**(2), 382–397.

Goldman, D. and Maestas, N. (2013), 'Medical expenditure risk and household portfolio choice', *Journal of Applied Econometrics* **28**(4), 527–550.

Gollier, C. (2001), *The Economics of Risk and Time*, MIT press.

Gollier, C. and Pratt, J. W. (1996), 'Risk vulnerability and the tempering effect of background risk', *Econometrica* pp. 1109–1123.

Graham, C. and Markowitz, J. (2011), 'Aspirations and happiness of potential latin american immigrants', *Journal of Social Research & Policy* **2**(2), 9.

Greene, W. (2003), 'Nonspherical disturbances—the generalized regression model', *Greene WH. Econometric Analysis. 5th edition. Upper Saddle River, NJ: Prentice Hall* pp. 191–214.

Greene, W. H. (2000), 'Econometric analysis 4th edition', *International edition, New Jersey: Prentice Hall* pp. 201–215.

Grossman, S. J. and Laroque, G. (1987), 'Asset pricing and optimal portfolio choice in the presence of illiquid durable consumption goods'.

Gu, B. (2014), 'Internal migration dominates population dynamics in china', *Asian Population Studies* **10**(1), 1–3.

Guedes Auditor, J. and Erlinghagen, M. (2021), The happy migrant? emigration and its impact on subjective well-being, *in* 'The Global Lives of German Migrants: Consequences of International Migration Across the Life Course', Springer International Publishing Cham, pp. 189–204.

Guiso, L., Haliassos, M. and Jappelli, T. (2002), *Household portfolios*, MIT press.

Guiso, L., Jappelli, T. and Terlizzese, D. (1996), 'Income risk, borrowing constraints, and portfolio choice', *The American Economic Review* pp. 158–172.

Guiso, L. and Paiella, M. (2008), 'Risk aversion, wealth, and background risk', *Journal of the European Economic Association* **6**(6), 1109–1150.

Haans, R. F., Pieters, C. and He, Z.-L. (2016), 'Thinking about u: Theorizing and testing u-and inverted u-shaped relationships in strategy research', *Strategic management journal* **37**(7), 1177–1195.

Hadjar, A. and Backes, S. (2013), 'Migration background and subjective well-being a multilevel analysis based on the european social survey', *Comparative Sociology* **12**(5), 645–676.

Haliassos, M. and Bertaut, C. C. (1995), 'Why do so few hold stocks?', *The Economic Journal* **105**(432), 1110–1129.

Hall, M. and Farkas, G. (2008), 'Does human capital raise earnings for immigrants in the low-skill labor market?', *Demography* **45**(3), 619–639.

Hamermesh, D. S. (1996), *Labor demand*, Princeton University Press.

Hanson, G. H. and Slaughter, M. J. (2002), 'Labor-market adjustment in open economies: Evidence from us states', *Journal of International Economics* **57**(1), 3–29.

Hausman, J., Stock, J. H. and Yogo, M. (2005), 'Asymptotic properties of the hahn–hausman test for weak-instruments', *Economics Letters* **89**(3), 333–342.

Heaton, J. and Lucas, D. (1997), 'Market frictions, savings behavior, and portfolio choice', *Macroeconomic Dynamics* **1**(1), 76–101.

Heaton, J. and Lucas, D. (2000), 'Portfolio choice in the presence of background risk', *The Economic Journal* **110**(460), 1–26.

Hendriks, M. and Bartram, D. (2019), 'Bringing happiness into the study of migration and its consequences: What, why, and how?', *Journal of Immigrant & Refugee Studies* **17**(3), 279–298.

Hering, L. and Poncet, S. (2010), 'Income per capita inequality in china: The role of economic geography and spatial interactions', *World economy* **33**(5), 655–679.

Hochguertel, S. (2003), 'Precautionary motives and portfolio decisions', *Journal of Applied Econometrics* **18**(1), 61–77.

Hu, J., Han, H., Zhang, Z. and Wu, X. (2023), 'Impact of homestead housing on the allocation of financial assets of chinese rural households', *Journal of the Asia Pacific Economy* **28**(1), 330–349.

Hu, X. (2012), 'China's' new generation'rural-urban migrants: Migration motivation and migration patterns', *Migration Information Source* .

Huang, Z. and Lu, Y. (2021), Analysis of asset risk and household financial asset allocation structure—empirical analysis from a nonlinear model, *in* 'E3S Web of Conferences', Vol. 235, EDP Sciences, p. 01039.

Hubbard, R. G., Skinner, J. and Zeldes, S. P. (1994), 'Expanding the life-cycle model: Precautionary saving and public policy', *The American Economic Review* **84**(2), 174–179.

Ioramashvili, C. (2023), 'It's not me, it's you: internal migration and local wages in great britain', *Regional Studies, Regional Science* **10**(1), 876–888.

Ivlevs, A. (2014), 'Happiness and the emigration decision', *IZA World of Labor* .

Ivlevs, A. et al. (2014), 'Happy moves? assessing the impact of subjective well-being on the emigration decision', *University of the West of England Economics Working Paper Series* **1402**, 1–23.

Jagannathan, R., Kocherlakota, N. R. et al. (1996), 'Why should older people invest less in stocks than younger people', *Federal Reserve Bank of Minneapolis Quarterly Review* **20**, 11–20.

Jeffrey, E. (1987), 'İntroduction: migration and health', *International Migration Review* **21**(3), v–xiv.

Jia, R. (2014), 'The legacies of forced freedom: China's treaty ports', *Review of Economics and Statistics* **96**(4), 596–608.

Jie, Y. and Mengyuan, J. (2018), Research on the influencing factors of household financial asset selection empirical evidence from china financial survey (chfs) data, *in* '4th International Conference on Economics, Management, Law and Education (EMLE 2018)', Atlantis Press, pp. 137–141.

Jin, L. (2016), 'Migration, relative deprivation, and psychological well-being in china', *American Behavioral Scientist* **60**(5-6), 750–770.

Jones, A. M. (1989), 'A double-hurdle model of cigarette consumption', *Journal of Applied Econometrics* **4**(1), 23–39.

Jones, A. M. and Yen, S. T. (2000), 'A box–cox double-hurdle model', *The Manchester School* **68**(2), 203–221.

Khan, S. U., Ramella, S., Rahman, H. U. and Hyder, Z. (2022), 'Household portfolio allocations: Evidence on risk preferences from the household, income, and labour dynamics in australia (hilda) survey using tobit models', *Journal of Risk and Financial Management* **15**(4), 161.

Knight, J. and Gunatilaka, R. (2010), 'Great expectations? the subjective well-being of rural–urban migrants in china', *World development* **38**(1), 113–124.

Knight, J., Song, L. and Huaibin, J. (2013), Chinese rural migrants in urban enterprises: Three perspectives, *in* 'The Workers' State Meets the Market', Routledge, pp. 73–104.

Knight, J. and Yueh, L. (2009), 'Segmentation or competition in china's urban labour market?', *Cambridge Journal of Economics* **33**(1), 79–94.

Kökönyei, G., Reinhardt, M., Pajkossy, P., Kiss, B. and Demetrovics, Z. (2008), 'Characteristics of personal strivings and their relationship with life satisfaction', *Cognition, Brain, Behavior* **12**(4), 409.

Koroļeva, I. (2022), A longitudinal study of factors affecting migrant subjective well-being: Empirical evidence from latvian emigrant surveys, *in* 'SOCIETY. INTEGRATION. EDUCATION. Proceedings of the International Scientific Conference', Vol. 2, pp. 9–24.

Lagakos, D., Marshall, S., Mobarak, A. M., Vernot, C. and Waugh, M. E. (2020), 'Migration costs and observational returns to migration in the developing world', *Journal of Monetary Economics* **113**, 138–154.

Lang, Y. and Wu, Y. (2014), An investigation of influencing factors on household portfolio in china—-based on micro econometric analysis of tobit model, *in* '2014 International Conference on Management, Education and Social Science (ICMESS 2014)', Atlantis Press, pp. 83–86.

Lapshyna, I. and Düvell, F. (2015), 'Migration, life satisfaction, return and development: the case of a deprived post-soviet country (ukraine)', *Migration and Development* **4**(2), 291–310.

Laurinavičius, A., Laurinavičius, A. and Laurinavičius, A. (2021), Macroeconomic variables influencing life satisfaction in the eu, and its impact on emigration, *in* 'International scientific conference" Contemporary issues in business, management and economics engineering 2021, 13–14 May 2021, Vilnius, Lithuania.', Vilnius Gediminas Technical University, pp. 1–12.

Leamer, E. E. and Levinsohn, J. (1995), 'International trade theory: the evidence', *Handbook of international economics* **3**, 1339–1394.

Lewis, E. (2011), 'Immigration, skill mix, and capital skill complementarity', *The Quarterly Journal of Economics* **126**(2), 1029–1069.

Lewis, E. and Peri, G. (2015), Immigration and the economy of cities and regions, *in* 'Handbook of regional and urban economics', Vol. 5, Elsevier, pp. 625–685.

Li, R., Wang, T. and Zhou, M. (2021), 'Entrepreneurship and household portfolio choice: Evidence from the china household finance survey', *Journal of Empirical Finance* **60**, 1–15.

Lin, T.-F. and Schmidt, P. (1984), 'A test of the tobit specification against an alternative suggested by cragg', *The Review of Economics and Statistics* pp. 174–177.

Lin, W.-c. and Lu, J.-r. (2012), 'Risky asset allocation and consumption rule in the presence of background risk and insurance markets', *Insurance: Mathematics and Economics* **50**(1), 150–158.

Love, D. A. and Smith, P. A. (2010), 'Does health affect portfolio choice?', *Health Economics* **19**(12), 1441–1460.

Luhmann, M., Lucas, R. E., Eid, M. and Diener, E. (2013), 'The prospective effect of life satisfaction on life events', *Social Psychological and Personality Science* **4**(1), 39–45.

Maddala, G. S. and Nelson, F. D. (1975), Specification errors in limited dependent variable models, Technical report, National Bureau of Economic Research.

Mara, I., Landesmann, M. et al. (2013), *Do I stay because I am happy or am I happy because I stay?: Life satisfaction in migration, and the decision to stay permanently, return and out-migrate*, Verein" Wiener Inst. für Internat. Wirtschaftsvergleiche"(WIIW).

Martínez-Espiñeira, R. (2006), 'A box-cox double-hurdle model of wildlife valuation: The citizen's perspective', *Ecological Economics* **58**(1), 192–208.

Matsukawa, I. (1991), 'Interregional gross migration and structural changes in local industries', *Environment and Planning A* **23**(5), 745–756.

McDowell, A. (2003), 'From the help desk: hurdle models', *The Stata Journal* **3**(2), 178–184.

McKenzie, D. and Rapoport, H. (2007), 'Network effects and the dynamics of migration and inequality: Theory and evidence from mexico', *Journal of development Economics* **84**(1), 1–24.

Melzer, S. M. (2011), 'Does migration make you happy? the influence of migration on subjective well-being.', *Journal of Social Research & Policy* **2**(2).

Michalos, A. C. and Maurine Kahlke, P. (2010), 'Stability and sensitivity in perceived quality of life measures: Some panel results', *Social Indicators Research* **98**(3), 403–434.

Michalos, A. C., Michalos, A. C. and Kahlke, P. M. (2017), 'Stability and sensitivity in perceived quality of life measures: Some panel results', *Development of Quality of Life Theory and Its Instruments: The Selected Works of Alex. C. Michalos* pp. 199–239.

Moffatt, P. G. (2005), 'Hurdle models of loan default', *Journal of the Operational Research Society* **56**(9), 1063–1071.

Mosselson, J. and Chinkondenji, P. (2020), 'Education, schooling, and migration', *Annual Review of Comparative and International Education 2019* pp. 323–334.

Mouw, T. (2016), 'The impact of immigration on the labor market outcomes of native workers: Evidence using longitudinal data from the lehd', *US Census Bureau Center for Economic Studies Paper No. CES-WP-16-56* .

Mrittika, S. and Marina-Selini, K. (2020), 'Migration and happiness: Evidence from germany', *Journal of Happiness Studies* **21**(8), 2931–2955.

Mutlu, S. and Gracia, A. (2006), 'Spanish food expenditure away from home (fafh): by type of meal', *Applied Economics* **38**(9), 1037–1047.

Nakazato, N., Schimmack, U. and Oishi, S. (2011), 'Effect of changes in living conditions on well-being: A prospective top–down bottom–up model', *Social indicators research* **100**, 115–135.

Neto, F. and Fonseca, A. C. M. (2016), 'The satisfaction with migration life scale', *International Journal of Intercultural Relations* **54**, 47–54.

Newman, C., Henchion, M. and Matthews, A. (2003), 'A double-hurdle model of irish household expenditure on prepared meals', *Applied Economics* **35**(9), 1053–1061.

Nowok, B., Van Ham, M., Findlay, A. M. and Gayle, V. (2013), 'Does migration make you happy? a longitudinal study of internal migration and subjective well-being', *Environment and Planning A* **45**(4), 986–1002.

Orrenius, P. M. and Zavodny, M. (2008), 'The effect of minimum wages on immigrants' employment and earnings', *ILR Review* **61**(4), 544–563.

Ortega, J. and Verdugo, G. (2011), 'Immigration and the occupational choice of natives: A factor proportions approach'.

Ossandon, G. and Castro-Santis, R. (2018), 'Effects of density-dependent migration on a population subjected to allee effect', *arXiv preprint arXiv:1809.05025* .

Otrachshenko, V. and Popova, O. (2012), 'Life (dis) satisfaction and the decision to migrate: Evidence from central and eastern europe', *SSRN* .

Ottaviano, G. I. and Peri, G. (2012), 'Rethinking the effect of immigration on wages', *Journal of the European Economic Association* **10**(1), 152–197.

Palia, D., Qi, Y. and Wu, Y. (2009), 'Heterogeneous background risks, portfolio choice, and asset returns: Evidence from micro-level data', *Concordia University and Rutgers Business School Working Paper* .

Palia, D., Qi, Y. and Wu, Y. (2014), 'Heterogeneous background risks and portfolio choice: Evidence from micro-level data', *Journal of Money, Credit and Banking* **46**(8), 1687–1720.

Palumbo, M. G. (1999), 'Uncertain medical expenses and precautionary saving near the end of the life cycle', *The Review of Economic Studies* **66**(2), 395–421.

Papastergiou, E., Latinopoulos, D., Evdou, M. and Kalogeresis, A. (2023), 'Exploring associations between subjective well-being and non-market values when used in the evaluation of urban green spaces: A scoping review', *Land* **12**(3), 700.

Pavot, W. and Diener, E. (1993), 'Review of the satisfaction with life scale.', *Psychological assessment* **5**(2), 164.

Pelizzon, L. and Weber, G. (2005), 'Efficient portfolios when housing is a hedge against rent risk', *unpublished paper, University of Venice and University of Padua* .

Peri, G. (2007), 'Immigrants' complementarities and native wages: Evidence from california'.

Peri, G. and Sparber, C. (2009), 'Task specialization, immigration, and wages', *American Economic Journal: Applied Economics* **1**(3), 135–169.

Peri, G. and Sparber, C. (2011), 'Highly educated immigrants and native occupational choice', *Industrial Relations: a journal of economy and society* **50**(3), 385–411.

Pratt, J. W. and Zeckhauser, R. J. (1987), 'Proper risk aversion', *Econometrica* **55**(1), 143–154.

Rachmat, N., Fanani, M., Darsono, D. and Suwarto, S. (2019), 'Factors affecting subjective well being of transfemoral prosthesis user in indonesia', *Advances in Social Sciences Research Journal* **6**(12), 269–274.

Rajani, N. B., Skianis, V. and Filippidis, F. T. (2019), 'Association of environmental and sociodemographic factors with life satisfaction in 27 european countries', *BMC public health* **19**, 1–8.

Ransom, T. (2016), 'The effect of business cycle fluctuations on migration decisions', *Available at SSRN 2741117* .

Renas, S. M. and Kumar, R. (1978), 'The cost of living, labor market opportunities, and the migration decision: A case of misspecification?', *The Annals of regional science* **12**, 95–104.

Roodman, D. (2009), 'Estimating fully observed recursive mixed-process models with cmp: working paper 168', *Center for Global Development, Washington, DC* .

Roodman, D. (2011), 'Fitting fully observed recursive mixed-process models with cmp', *The Stata Journal* **11**(2), 159–206.

Roodman, D. (2023), 'Cmp: Stata module to implement conditional (recursive) mixed process estimator'.

Rosen, H. S. and Wu, S. (2004), 'Portfolio choice and health status', *Journal of Financial Economics* **72**(3), 457–484.

Samuel, R. and Hadjar, A. (2016), 'How welfare-state regimes shape subjective well-being across europe', *Social Indicators Research* **129**, 565–587.

Schiele, M. (2021), 'Life satisfaction and return migration: analysing the role of life satisfaction for migrant return intentions in germany', *Journal of ethnic and migration studies* **47**(1), 110–129.

Selezneva, E. (2011), 'Surveying transitional experience and subjective well-being: Income, work, family', *Economic systems* **35**(2), 139–157.

Sen, A. (1990), 'Development as capability expansion', *The Community Development Reader* **41**, 58.

Siebenmorgen, N. and Weber, M. (2003), 'A behavioral model for asset allocation', *Financial Markets and Portfolio Management* **17**(1), 15.

Silva, J., De Keulenaer, F. and Johnstone, N. (2012), 'Environmental quality and life satisfaction: Evidence based on micro-data'.

Sinai, T. and Souleles, N. S. (2005), 'Owner-occupied housing as a hedge against rent risk', *The Quarterly Journal of Economics* **120**(2), 763–789.

Sjaastad, L. A. (1962), 'The costs and returns of human migration', *Journal of political Economy* **70**(5, Part 2), 80–93.

Smith, M. D. (2003), 'On dependency in double-hurdle models', *Statistical Papers* **44**, 581–595.

Stark, O. and Yitzhaki, S. (1988), 'Labour migration as a response to relative deprivation', *Journal of Population Economics* **1**(1), 57–70.

Stock, J. and Yogo, M. (2005), 'Asymptotic distributions of instrumental variables statistics with many instruments', *Identification and inference for econometric models: Essays in honor of Thomas Rothenberg* **6**, 109–120.

Storesletten, K., Telmer, C. I. and Yaron, A. (2004), 'Cyclical dynamics in idiosyncratic labor market risk', *Journal of Political Economy* **112**(3), 695–717.

Su, S.-J. B. and Yen, S. T. (1996), 'Microeconometric models of infrequently purchased goods: an application to household pork consumption', *Empirical Economics* **21**, 513–533.

Sullivan, P. (2010), 'Empirical evidence on occupation and industry specific human capital', *Labour economics* **17**(3), 567–580.

Switek, M. (2016), 'Internal migration and life satisfaction: Well-being paths of young adult migrants', *Social Indicators Research* **125**, 191–241.

Tan, S., Li, Y., Song, Y., Luo, X., Zhou, M., Zhang, L. and Kuang, B. (2017), 'Influence factors on settlement intention for floating population in urban area: A china study', *Quality & Quantity* **51**, 147–176.

Vearey, J., Hui, C. and Wickramage, K. (2020), 'Migration and health: current issues, governance and knowledge gaps', *World migration report* **2020**(1), e00017.

Veenhoven, R. (1995), 'World database of happiness', *Social Indicators Research* **34**, 299–313.

Veenhoven, R. (2011), Happiness: Also known as "life satisfaction" and "subjective well-being", *in* 'Handbook of social indicators and quality of life research', Springer, pp. 63–77.

Viceira, L. M. (2001), 'Optimal portfolio choice for long-horizon investors with nontradable labor income', *The Journal of Finance* **56**(2), 433–470.

Vincenty, T. (1975), 'Direct and inverse solutions of geodesics on the ellipsoid with application of nested equations', *Survey review* **23**(176), 88–93.

Wang, A. (2009), 'Interplay of investors' financial knowledge and risk taking', *The Journal of Behavioral Finance* **10**(4), 204–213.

Wang, H. (2016*a*), *Private Market Alternatives for Maintaining Wetland Viability in Coastal Louisiana: A Double-Hurdle Approach*, Louisiana State University and Agricultural & Mechanical College.

Wang, N. (2016*b*), 'Changing trend of population migration in china and its spatial pattern', *Chinese Journal of Urban and Environmental Studies* **4**(04), 1650032.

Wener, R. E. and Evans, G. W. (2011), 'Comparing stress of car and train commuters', *Transportation Research Part F: Traffic Psychology and Behaviour* **14**(2), 111–116.

West, T. and Worthington, A. C. (2017), 'The impact of major life events on australian household financial decision-making and portfolio rebalancing', *Available at SSRN 2908595* .

Woodridge, J. (2010), 'Econometric analysis of cross section and panel data, 2nd edtion'.

Wright, K. (2011), 'Constructing migrant wellbeing: An exploration of life satisfaction amongst peruvian migrants in london', *Journal of Ethnic and Migration Studies* **37**(9), 1459–1475.

Wu, W. and Gaubatz, P. (2013), *The Chinese City*, 2nd edn, Routledge.

Xuguang, S. and Jiajia, H. (2022), 'The impact of migration on the wage of locals: Based on the perspective of labor skill divergence', *Journal of Finance and Economics* **48**(06), 19–33.

Yang, J., Liu, K. and Zhang, Y. (2019), 'Happiness inequality in china', *Journal of Happiness Studies* **20**, 2747–2771.

Yang, S., Chen, L. and Bi, X. (2023), 'Overtime work, job autonomy, and employees' subjective well-being: Evidence from china', *Frontiers in Public Health* **11**, 1077177.

Yang, S., Shu, T. and Yu, T. (2022), 'Migration for better jobs or better living: Shifts in china', *International Journal of Environmental Research and Public Health* **19**(21), 14576.

Yen, S. T. (1993), 'Working wives and food away from home: The box-cox double hurdle model', *American Journal of Agricultural Economics* **75**(4), 884–895.

Yen, S. T. and Jensen, H. H. (1996), 'Determinants of household expenditures on alcohol', *Journal of Consumer Affairs* **30**(1), 48–67.

Yen, S. T. and Su, S.-J. (1995), 'Modeling us butter consumption with zero observations', *Agricultural and Resource Economics Review* **24**(1), 47–55.

Zhang, Y., Jia, Q. and Chen, C. (2021), 'Risk attitude, financial literacy and household consumption: Evidence from stock market crash in china', *Economic Modelling* **94**, 995–1006.

Zhao, H., Chang, J. and Wang, J. (2023), 'What pulls chinese migrant workers back to the countryside? an analysis from a family concerns perspective', *International Review of Economics & Finance* **84**, 801–812.

Zhao, X. (2020), 'Migrants and urban wage: Evidence from china's internal migration', *China Economic Review* **61**, 101287.

Zheng, Z. and Yang, G. (2016), 'Internal migration in china: Changes and trends', *Contemporary demographic transformations in China, India and Indonesia* pp. 223–237.