HCEO WORKING PAPER SERIES

Where Are Migrants From? Inter- vs. Intra-Provincial Rural-Urban Migration in China

Yaqin Su
Petros Tesfazion
Zhong Zhao

Working Paper

2018-003

01/2018

HUMAN CAPITAL AND
ECONOMIC OPPORTUNITY
GLOBAL WORKING GROUP

The University of Chicago
1126 E. 59th Street Box 107
Chicago IL 60637

www.hceconomics.org

Where Are Migrants from?
Inter- vs. Intra-Provincial Rural-Urban Migration in China[†]

Yaqin Su

Center for Economics, Finance and Management Studies

Hunan University, China

yaqinsu@hnu.edu.cn

Petros Tesfazion

Central College, Pella, Iowa

ptesfazion@yahoo.com

Zhong Zhao[*]

School of Labor and Human Resources

Renmin University of China

mr.zhong.zhao@gmail.com

Abstract

Using a representative sample of rural migrants in cities, this paper investigates where the migrants in urban China come from, paying close attention to intra-provincial vs. inter-provincial migrants, and examining the differences in their personal attributes. We find that migrants who have come within the province differ significantly from those who have come from outside of the province. Using a nested logit model, we find that overall, higher wage differentials, larger population size, higher GDP per capita, and faster employment growth rate are the attributes of a city that attract migrants from both within and outside province. In addition, moving beyond one's home province has a strong deterrent effect on migration, analogous to the "border effect" identified in international migration studies. We also explore the role of culture, institutional barrier, and dialect in explaining such a pronounced "border effect".

Keywords: Rural-urban migration; Inter- vs. intra-provincial migration; Border effect; China

JEL Classification: J62; O15

[†] We are very grateful for comments from Belton M. Fleisher (editor), Xiaojun Wang (guest editor) and an anonymous referee. The collection of the Rural Urban Migration in China (RUMiC) data used in this paper is financed by IZA, ARC/AusAid, the Ford Foundation, and the Ministry of Labor and Social Security of China. Zhong Zhao would like to acknowledge financial support by Renmin University of China: the special developing and guiding fund for building world-class universities (disciplines) (Project No. 16XNL005).

[*] Corresponding author.

1. Introduction

China has witnessed an extraordinary internal migration unparalleled in world history. Since 1979, over 500 million people have been added to China's urban population, of which 78 percent was attributable to rural-to-urban migrants. As China moves away from its export-driven growth to focus on stimulating domestic demand, the consumption and investment potentials brought by migrants will play a critical role in sustaining the long term economic growth of Chinese cities. Meanwhile, the Chinese government has pledged to gradually relax the household registration system, known as *hukou*, in an attempt to encourage labor mobility and stimulate spending in cities.

Based on the latest survey of the National Bureau of Statistics (NBS), in 2015 there are a total of 277 million rural workers in China, among which 168 million are rural migrant workers who have left the village or town associated with their *hukou* and sought employment elsewhere. As indicated in Figure 1, the number of rural migrant workers has grown from 104 million in 2002 to 168 million in 2015, despite the fact that the rate of growth has declined sharply since 2010. In 2015, 45.9% of rural migrants are inter-provincial migrants, while 54.1% are intra-provincial migrants. The share of migrants who choose to work at prefectural-level cities or above is 80% for inter-provincial migrants, and 54.6% for intra-provincial migrants.

Rural-urban migration in China has been intensely studied from various perspectives over the past two decades, primarily due to its inextricable link with China's ongoing urbanization process and its profound social ramifications. For examples, there are many studies that examine the discrimination against rural migrants and their inferior social and labor market positions (e.g., Chan and Buckingham, 2008; Chan and Zhang 1999). The main consensus is that the majority of rural migrants in cities lead second-class lives without much access to urban benefits. Studies have also suggested negative health and wellbeing consequences of rural-urban migration to migrants themselves (Song and Sun, 2015), and to the elderly and children left behind in the rural areas (Ao, Jiang, and Zhao, 2016; Li, Liu, and Zang, 2015; Xu and Xie, 2015; Murphy, Zhou, and Tao, 2015; Logan, 2011; De Brauw and Mu, 2011; Lee, 2011).

An important strand of research is on rural-urban migration decision and determinants, e.g. Zhao (1999a, b), Zhu (2002) and Cheng et al (2006).[1] Although

[1] We will review related literature in next section.

previous research has generated an impressive body of work, most of the existing studies are from the perspective of migrant sending area, and there are few studies from the perspective of the urban area, e.g., where do the migrants come from? Are the inter- and intra-provincial migrants different? Whether the cities are still able to attract cheap labor from the rural area? Which city attributes are attractive to migrants?

These questions are important in several aspects. First, it is generally acknowledged that rural-urban migrants have greatly contributed to China's economic growth by providing inexpensive labor to the manufacturing and service industries that have been rapidly expanding in urban China. However, since 2004, there has been mounting concerns regarding whether China has run out of the cheap surplus of rural workers, which could potentially curb the country's future economic growth. Many studies claim that China has reached a "Lewis turning point" in economic development, demonstrated by rising wages in urban areas and the exhaustion of rural surplus labor (e.g., Cai and Wang, 2010; Garnaut and Song, 2013). In this context, examining which city attributes are attractive to a migrant is particularly relevant and meaningful. Second, the sources of migrants play an important role in shaping the spatial distribution of migrant populations. According to the NBS, the patterns of rural-urban migration have shown several changes in recent years. There has been a sudden drop in the proportion of rural workers undertaking inter-provincial migration recently: in 2011, intra-provincial migrants out-numbered inter-provincial ones. Meanwhile, the proportion of migrants flowing to the eastern region has shown a continuous decline. This can be due to a reduced rural-urban income gap; thus, more migrants prefer to find employment in close proximity of their rural hometown, which could further contribute to the economic development in rural areas. Third, migrant workers typically take the jobs at the lowest end of the earning spectrum and jobs shunned by local workers. Based on the Rural Urban Migration in China (RUMiC) data, a higher proportion of inter-provincial migrants are engaged in service jobs; more specifically, a larger share of them work in personal care jobs such as beauty services, barber, etc., whereas a smaller proportion work in restaurants. As the proportion of inter-provincial migration shrinks, the price associated with these services will increase, causing a higher living cost in cities. In sum, the research questions we aim to address in this paper are important from both urban and rural development perspectives; and are highly relevant for the welfare of both urban residents and rural migrants.

In this paper, we use a unique dataset that covers rural-urban migrant workers in 15 major migrant receiving cities in China; we aim to shed some light on rural-urban migration from an urban perspective, and try to answer the above questions. While many Chinese studies have focused on *why* a person migrates, this paper focuses on *where* a migrant comes from and *what* city attributes attract them. This paper distinguishes itself from the existing research in several aspects.

First, it pays special attention to the composition of migrants in a destination city, and investigates the determinants of a migrant's choice regarding moving within versus outside of his province of origin. We find strong evidence that personal attributes play an important role in this regard, a fact typically overlooked by studies using aggregated provincial data. Our empirical results suggest that migrants who have come from other provinces differ significantly from the migrants from the same province as the destination city in a variety of individual characteristics. In general, the migrants from other provinces tend to be younger, more likely to be male, usually in better physical condition, but have fewer years of formal education. In addition, we find that moving beyond one's home province has a strong deterrent effect on migration, analogous to the "border effect" identified in international migration studies; this finding has important policy implications in the sense that if a city wants to attract migrants, it will be cheaper and easier to attract migrants from the home province.

Second, we explore the role of city attributes in attracting migrants, and find that overall, higher wage differentials, larger population size, higher GDP per capita, and faster employment growth rate are the attributes of a city that attract migrants from both within and outside province. Meanwhile, distance between a city and migrants sending area has a significant deterrent effect, although this effect attenuates as distance increases.

Third, this paper employs an increasingly popular discrete choice model—the nested logit model. Thus far the majority of studies that examined the determinants of migration in China use either logit or multinomial logit models. However, the validity of multinomial logit model rely on the independence from irrelevant alternatives (IIA) assumption, that is, the odds ratio for any pair of alternatives will stay the same, regardless of the changes in the attributes of other alternatives, which is clearly questionable in some research scenarios. Compared with the multinomial logit model that requires IIA assumption, the nested logit model allows the alternatives to be correlated due to unobserved characteristics. That is, the set of alternatives can be allocated into nests based on perceived similarities. This methodology is more flexible and has not been widely utilized in the research on rural-urban migration in China.

The remainder of the paper is organized as follows. Section 2 provides the background and discusses related empirical studies on rural-to-urban migration in China. Section 3 describes the data and descriptive statistics. Section 4 discusses the estimation techniques and presents empirical results, and the last section concludes.

2. Background and Literature review

Prior to the 1980s, rural-urban migration in China was minimal. The imposition of strict state control over food, clothes, jobs, medical care and housing denied rural residents the basic necessities needed to survive in urban environments. The first wave of rural-urban migrants in the early 1980s was analyzed in some papers, e.g. Chen (2012). In his paper, Chen demonstrates that the Household Responsibility System (HRS) implemented in 1980, which resulted in changes in the property rights and abolished the tight monitoring of labor allocations in rural households, led directly to the emergence of rural-urban non-*hukou* migration in the early reform period. Many rural residents began to search for jobs in nearby cities and the number of non-*hukou* intra-provincial rural migrants increased dramatically. In his analysis, only intra-provincial migration is considered since inter-provincial migration barely existed in the early stage of China's internal migration.

From 1982 to 1989, the total stock of "floating population" (*liudong renkou*) increased from 11 million to about 18 million (Liang, 2001).[2] Using the 1987 1% population survey data, Ma, Liaw and Zeng (1997) pointed out that although the volume of migration during the mid-1980s is not trivial, the rate of migration is extremely low--roughly 0.7% per year, compared to Japan's 5.39%, Peru's 6.5% inter-provincial migration rate (thus indicating a much higher overall migration rate), and Thailand's 4.1% during the mid-1980s.

[2] "Floating population" is probably the most easily available and consistent measure of migration based on various sources of data. It's defined as the migrants who are without local *hukou* and have stayed in the current location for more than 6 months (or a year depending on the census year). Note that it is not equivalent to "rural migrant labor" that we focus on in this paper since the floating population also contains those with an urban *hukou*, yet have stayed in a different location for a significant length of time.

Temporary migration in China began to accelerate in the late 1980s: in a single year from 1989 to 1990, China's floating population increased by 7 million. And the momentum continued through the 1990s (Liang, 2001). In the 1990s, the rapid expansion of the export-oriented sectors in China's eastern coastal regions created a vigorous demand for unskilled labor, and rural laborers started to pour into the eastern regions, especially Guangdong province, in search of job opportunities. The massive flow of relatively young workers migrating from China's interior to its coastal regions has drawn most of the attention from researchers, mainly due to its undeniable role in China's export-led growth over the past three decades.

However, inter-provincial migration only provides a partial picture of the overall migration pattern. Surprisingly, intra-provincial migration has received far too little attention in the existing literature in spite of the fact that various data sources indicate that intra-provincial migration has dominated its inter-provincial counterpart in terms of volume and scope for the most part of China's rural-urban migration history since 1980. Intra-provincial migration plays an equally, if not more important role, in determining the spatial distribution of migrant populations. Based on China's national population censuses, the share of inter-provincial migrants in the total floating population is 29.4%, 25.5% and 38.9% for each five-year period ending in 2000, 2005 and 2010, respectively (Taylor, 2011; Liang, Li and Ma, 2014). It is worth pointing out that different geographical boundaries are used to define a migrant in census data of different years, which makes it difficult to compare the rate of inter-provincial migration over time. The earlier censuses, i.e., the 1990 and 1995 censuses, used *county* as the geographic boundary that a migrant needed to cross to qualify as a migrant. Intra-county migration, which representing a significant portion of overall migration, was ignored in the 1990 and 1995 censuses. Thus, combining the data from the 1990, 2000, and 2010 censuses, along with the inter-censal 1% population surveys conducted in 1995, 2005 and 2015, the proportion of inter-provincial migrants as a share of *inter-county* migrants are more prominent: 27.7%, 32.1%, 53.9%, 55%, 50.3%, 45.9% corresponding to each five-year period ending in 1990, 1995, 2000, 2005, 2010, 2015, respectively (Chan, 1994; Liang, 2001; Liang, 2004, Liang, Li and Ma, 2014).

No doubt, inter-provincial migration has grown at a faster rate than its intra-provincial counterpart since the 1990s. Based on the 1995 census, inter-provincial migrants increased by 68% from 1987 to 1995, while the number of intra-provincial migrants changed only slightly over the same period (Liang, 2001). Many researchers have ascribed this phenomenon to an increasing regional inequality. As a result, in the late 2000s, the inter-provincial rural migrants seemed to have outpaced their intra-provincial counterparts and become the dominant form of rural migrant labor.

According to the surveys of the NBS over the period 2008－2015, several noticeable changes have taken place in terms of the size, rate of growth, and spatial patterns of migrant workers.[3] First, the gross stock of rural workers (including who migrate to cities and who work locally) has increased from 225 million in 2008 to 277 million in 2015, whereas the rate of growth has demonstrated a clear downward trend over the years.[4] Second, as indicated in Figure 2, in 2011, intra-provincial rural migrant workers once again outnumbered inter-provincial ones, becoming the dominant form of rural migrants. Third, the proportion of migrant workers flowing to the eastern region has shown a continuous decline, whereas the proportion that goes to the mid and western regions has been increasing over the period. Similarly, the proportion of migrant workers that flow to the Yangtze River Delta and Pearl River Delta has also shown a continuous decline.

Following, we briefly discuss some of the most relevant works with an emphasis on the evolution of the spatial patterns of migration in China. To provide a broad context concerning migration studies in general, the existing theoretical models in the migration literature can be broadly classified into three groups: First, the Lewis dual-sector models in the 1950s and 1960s that focus on rural surplus labor as the main driver of labor migration; second, the Harris-Todaro models developed in the 1970s and 1980s that focus on the role of rural-urban migration in urban unemployment; and third, a series of microeconomic models, also referred to as New Economic of Labor Migration (NELM), which focus on households as the decision-making unit, migrant network effect and the impact of migration on rural development.

Prior to the 1990s, the body of research on internal migration in China was quite small, mainly due to the lack of data. Information on migration was not included for the first three population censuses in 1952, 1964 and 1982. Wu and Zhou (1996) state that

[3] Since 2008, the National Bureau of Statistics (NBS) has conducted surveys on rural workers to track their magnitude, direction of flow, structure, employment, income and expenditures, housing, social security, etc.

[4] We need to differentiate rural workers from rural *migrant* workers since rural workers are those with rural *hukou* and choose to work in a non-agricultural job, which includes both migrants and non-migrants.

migration research in China was still at an infant stage at that time with most studies confined to qualitative or descriptive discussion. But with the rapid surge of rural-urban migration in China, research concerning internal migration has proliferated in the past two decades, leading to a body of valuable empirical findings.

In terms of determinants of migration decisions, there are many factors that contribute to decision of "why migrate?" This decision is commonly conceptualized as being influenced by "push factors"-- for example, poverty or lack of job opportunities in rural origins, coupled with "pull factors"-- such as jobs, amenities, and public goods in the destination localities. Using the 1995 rural household survey from Sichuan province, Zhao (1999a) finds that migration decisions are largely determined by the "push factors"-- unavailability of land and employment opportunities in the rural areas. Nevertheless, rural people appear to have a strong preference to remain in rural areas. Zhao (1999b) further stresses the importance of the shortage of farmland and the abundance of rural labor as the most important determinant of labor migration. In addition, she finds that rural taxation exerts a significant "push" effect on the migration decision. By observing migrant households' spending behaviors, she infers that most rural migrants will only stay in cities temporarily since the earnings from migration have little impact on the consumption patterns of rural households. Echoing the importance of "push factors" in the migration decision, Cheng et al. (2006) state that the backwardness of the industrial technology in rural areas is the key reason why many rural workers leave rural areas in search of non-farm work elsewhere. Using the 1993 survey data in Hubei province, Zhu (2002) presents evidence that the rural-urban income gap is the primary factor underlying rural-urban migration. Seeborg et al. (2000) examine the causes and consequences of the large "floating population" in China and state that a large surplus of rural labor supply, combined with a surge in demand for rural workers in urban industrial sectors, jointly determines the rural-urban labor migration in China. Further, they emphasize that the development of a contract labor system and the emergence of a private sector also substantially contributed to the massive internal migration in China.

Thus far the majority of studies that examined the migration decision in China rely on either binary choice (logit or probit) or multinomial choice (mutinomial logit) models. Based on conditional logit analysis, Xia and Lu (2015) focus on how public goods, particularly basic education and medical services, affect the location choice of inter-city migrant workers in China. Using China's national population survey conducted in 1987, Ma and Liaw (1997) applied a two-level nested logit model to examine the destination choice of Chinese young migrants aged 17-29. To address the endogeneity issue that normally arises in the context of migration decisions, some researchers resort to simultaneous estimation with a Heckman's two-step procedure (e.g., Meng, 2001) or a switching regression with selection bias correction (Zhu, 2002). Jalan and Ravallion (2000) use a quantile regression to examine the effect of income risk on migration decisions. The ordinary least square (OLS) and the weighted least square (WLS) estimates are widely used in studies that examine the province-to-province migration flows. It is worth pointing out that many Chinese studies used inter-provincial migration flows drawn from the population census data which exclude the migrants who have moved within their home provinces and treat them as "non-movers" (Bao, Hou, and Zhao, 2007; Lin, Wang and Zhao 2004; Poncet 2006). In the recent decade, more and more researchers prefer to use micro data to examine migration behaviors. There are obvious advantages associated with using micro data instead of provincial aggregates. First and foremost, the use of individual migration choice can help alleviate reverse causality issues. Second, it allows us to examine how individual heterogeneity, such as age, gender, education, etc., affects migration outcomes.

From a policy perspective, the vast majorities of studies consider restriction on migration a distortion of the labor market and thus should be gradually abolished. For example, Zhao (1999a) disapproves of the artificial barriers caused by the *hukou* system since they impede the flow of rural surplus labor to more productive uses, resulting in tremendous efficiency and productivity losses in the economy. Au and Henderson (2006) assert that restrictions on labor mobility lead to an insufficient amount of agglomeration of economic activities in both rural and urban sectors, incurring large GDP losses and undermining the urbanization process in China. Similarly, Chang and Brada (2006) observe that China's urbanization process slowed in spite of massive migration from rural to urban areas, partly due to the fact that most migrants are temporary in nature.

The institutional barriers that keep rural migrants as outsiders in the urban society result in substantial losses in employment and GDP. Han (2007) explicitly contends that the Chinese government should adopt favorable policies for rural migrants to facilitate balanced development between urban and rural areas and promote equality, humanity and economic prosperity.

3. Data and Descriptive Statistics

3.1 Data

Analyses in this paper are mainly based on the Rural Urban Migration in China (RUMiC) survey data. The RUMiC survey is a large-scale household survey initiated by a group of researchers at the Australian National University, the University of Queensland and the Beijing Normal University, and supported by the Institute for the Study of Labor (IZA). The surveys consist of 5,000 migrant households in cities, 5,000 urban residence households and 8,000 rural households (with or without migrant workers). The survey is designed to provide a longitudinal dataset covering a four-year span from 2008 to 2012. Since 2008, five waves of the migrant household surveys have been conducted. At this point, the first two waves' survey data are made available to researchers in general.

The RUMIC survey selects the locations based on whether a province is one of the major sending or receiving regions. As a result, the migrant sample was conducted in 15 cities across nine provinces or metropolitan areas: Shanghai, Guangdong, Jiangsu, Zhejiang, Anhui, Hubei, Sichuan, Chongqing and Henan, where the first four locations are the largest migration destinations and the remaining five are the among largest migration sending areas.

The RUMiC survey provides a representative sample of rural migrants in the cities, including those who live at their workplaces such as factory dormitories and construction sites, and this is the sample used in this paper. So far, most existing surveys regarding rural migrants use registered residential addresses as a basis for sampling. The biases associated with such a sampling frame can arise from two sources. First, although China requires anyone who stays more than 3 days in a place other than where his/her *hukou* is registered to report to the local police and obtain a "temporary residence permit", more often than not, many rural migrants fail to comply with this regulation to avoid the fee involved in the process. As a result, a large proportion of temporary migrant workers was left out of the sample and became invisible in most censuses carried out at the national level. Secondly, surveys based on the residential addresses leave out a non-trivial proportion of migrants who live in dormitories and construction sites provided by their employers. Based on the RUMiC survey, 41 percent of rural migrants live in the dormitory provided by their workplace; 8 percent of households live at the construction sites or other types of working area; only half of the households rent their housing independently or share it with other people. To address these sampling biases, the RUMiC survey employed a workplace-based sampling strategy, thereby providing a more accurate representation of rural temporary migrants.[5]

This paper mainly utilizes the Migrant Survey of the first two waves in 2008 and 2009. Despite substantial efforts made by the survey team to track every migrant as long as they remain in the surveyed cities and villages, there is a high attrition rate in the migrant sample from 2008 to 2009. By 2009, the RUMiC survey had lost track of more than half (58 percent) of the migrant individuals contained in the 2008 migrant sample. This is partly due to the fact that migrant workers are temporary in nature and highly mobile; they move from city to city in search of higher paying jobs. The high attrition rate can also be largely attributed to the global financial crisis that hit China in 2009, which especially affected the export industry in coastal cities such as Guangzhou and Dongguan, where most migrant workers are concentrated. About 23 million migrant workers lost their jobs in early 2009. To restore the migrant sample to the original size, the RUMiC team surveyed a new random sample of migrants in 2009. As a result, the dataset utilized in this paper contains 8,449 individual migrants based on the survey in 2008 and 5,426 additional migrant individuals in 2009. The resulting sample has 13,872 individual records. The panel dimension offered by the RUMiC data is not utilized in this analysis.

5 As the first step, each selected city is divided into hundreds of equal-sized blocks within defined city boundaries (i.e., 0.5 km by 0.5 km), from which 20-50 blocks are randomly drawn for survey purposes. Within each selected block, a census of workplaces is undertaken and information about the number of migrant workers at each workplace is collected. The total number of migrant workers can then be aggregated and used as an estimate for the total size of the migrant population in the cities. Finally, a random sample of migrant workers is selected to have a face-to-face interview with the enumerator. For detailed information on sampling design including methodology and implementation manuals, see Gong, Kong, and Meng (2008): “Rural-urban migrants: a driving force for growth”, in China’s Dilemma.

3.2 Descriptive statistics

Table (1) presents basic characteristics of rural migrants based on 12,322 migrant individuals aged 16 to 64 in the cities. The migrants in our sample are, on average, 31 years old. 66% of migrants are young workers aged 16 to 35, 57% are male, consistent with the conventional view that men are more likely to migrate than women. Their average educational attainment is 9.2 years, only slightly higher than the 9 years of education mandated by the Compulsory Schooling Law in China. In addition, the majority of them are married (61 percent) and have one child. On average, they make around 1615 yuan (about 250 U.S. dollars) per month in the cities. The self-employed ones earn noticeably higher income than the wage earners – 872 yuan (around $140) more per month on average. The self-employed migrants are not a trivial portion, constituting 22% of the entire migrant sample. On average, eight years have elapsed since they first migrated from their home villages, and most of them have changed cities in between. Overall, they are very healthy: 84% of them consider themselves in good or excellent health condition.

The RUMiC data indicate that overall 55% of migrants come from home provinces of the migrants. Table (2) presents the source of migrants working in the 15 cities in our sample. Although the rural migrants in our dataset come from all provinces in the China, the eight provinces included in row 1 of Table (2) provide over 85% of the migrant workers for the 15 cities. A clear pattern emerges from Table (2): the majority of migrant workers are from the rural areas of the home province of the migrants. Take Chengdu－the capital city of Sichuan province, as an example, 62% of migrant workers are from Sichuan, and another 4% from Chongqing, a municipality in the proximity of Sichuan. The share of migrants from elsewhere is less than 34%. For the case of Guangzhou, 68% of the migrant workers come from rural areas of Guangdong, only 2% from Sichuan, another 2% from Hubei, and an additional 3% from Henan, together accounting for 75% of rural migrant workers in Guangzhou. The source of migrants for other cities shares a similar pattern.

To see if the intra-provincial migrants differ from their inter-provincial counterparts in individual characteristics, we performed *t*-tests on the difference in means for these two groups using both RUMiC data and the 2005 1% population census. The results are summarized in Table (3). Based on RUMiC data, with the exception of no discernible difference in marital status and degree of risk aversion, these two groups are significantly different in terms of all other personal characteristics. In summary, migrants from other provinces are younger, more likely to be male, less educated and having better health. Besides, they are more mobile, i.e. they have migrated for a shorter period of time on average yet changed more cities than their intra-provincial counterpart. Focusing on the means through which the migrants find a job, intra-provincial migrants are more likely to acquire jobs through social network, such as family, relative and friends. Furthermore, inter-provincial migrants are more likely to take jobs that provide more job security, i.e., jobs that provide unemployment insurance, pension and injury insurance; inter-provincial migrants consist of a higher proportion of contracted workers, fewer temporary workers and family helpers, and, they are also less likely to be self-employed than intra-provincial migrants. The comparisons based on 2005 1% population census yield consistent result, indicating that the comparisons based on the RUMiC data are convincing.

4. Empirical Implementation

4.1 Basic setup

To model where the migrants come from, and take into account the role of personal characteristics of migrants and city attributes, we rely on a random utility framework that incorporates personal attributes of the migrant, the economic condition at the alternative destination cities, and the costs of moving such as distance. Suppose that a migrant *i* faces J alternatives, indexed by *j* = 1, 2, ..., J, the utility he derives from choosing *j*th alternative is modeled as:

$$U_{ij} = V_{ij} + \varepsilon_{ij} \qquad (1)$$

Assume that each migrant is rational and utility-maximizing, the probability of a migrant choosing alternative *j* equals the probability of such choice yielding him the maximum level of utility. Let the deterministic component of the utility function for individual *i*, V_{ij}, be a function of personal and location-specific characteristics,

$$V_{ij} = F(X_i, Y_j, Z_{ij}) \qquad (2)$$

where X_i is a vector of individual characteristics such as age, sex, marital status and education; Y_j is a set of destination-specific attributes that do not depend on one's origin. The variables identified to be important factors in the existing literature include the economic strengths of a locality (such as GDP and average wage), population size, employment growth, cost of living such as rent, amenities, etc.; and Z_{ij} is a set of variables that depend on both origin and destination characteristics. For example, the distance between the origin and destination and wage differentials between the destination city and rural village. Note that although distance is usually pinpointed as one of the most important deterrence to migration, the cost of migration goes beyond what is captured by distance alone. There are other “noneconomic factors” that impose costs on migration, to name a few, “psychic costs” associated with separation from family, safety concerns, lack of access to urban public services especially basic education to migrant’s children, not to mention the deep-rooted discrimination against rural workers from urban society.

We employ a two-level nested logit model to examine the migration decision-making process as illustrated in Figure 3. The upper-level branches represent the choices of whether to migrate within the source province or migrate to a province other than that of one’s origin. It seems reasonable to assume that personal attributes such as age, gender, health and education level will exert a bigger influence on the top-level decisions. Conditional on the decision made at the top-level model, the bottom-level choice set is partitioned into the city destinations within one’s origin province and those outside one’s origin. For example, if the rural migrants from Hubei province choose to stay within Hubei, the city choice will be Wuhan -- the capital city of Hubei. On the other hand, if they choose to move inter-provincially, the set of city choices will be the remaining 14 cities excluding Wuhan. Note that as in the case of Hubei province, the “Within” branch has a degenerate structure since there is only one choice in the nest of “Within”, and a non-degenerate structure exists for the “Out” branch, which contains all cities in the non-origin provinces.

Note that although the nested logit tree structure is the same for all migrants from the same origin, it varies for migrants from different provinces. This constitutes a major challenge for estimating a nested logit model based on the inter- and intra-provincial choice on the top-level. For the nested logit model to work, many standard statistical software packages require the set of alternatives to be partitioned into non-overlapping subsets. In other words, an alternative needs to be assigned to only one nest. To address this limitation, we opt to estimate the nested logit model sequentially by estimating the conditional probabilities first (bottom-level), followed by estimating the unconditional probabilities (top-level). Ideally, the parameters in the model are estimated simultaneously by MLE (maximum likelihood estimation), which utilizes all the information available in the data. Nevertheless, sequential estimation provides consistent estimates when the simultaneous estimation is difficult to implement. Note that, although the sequential structure of the nest logit model is often interpreted as implying higher-level decisions are made first, followed by decisions at lower levels, a temporal ordering is not necessarily implied by the nest logit model. In fact, the nested logit model is appropriate as long as we believe the alternatives are similar to each other in unobserved factors.

4.2 Lower-level conditional probabilities

Since the nested logit model specified in this paper does not satisfy the requirement that each alternative belongs to only one nest. We use sequential estimation techniques that provide consistent estimates for such a model. The lower-level conditional probabilities take the form:

$$P_{ij|w} = \frac{\exp(\beta' Y_j + \gamma Z_{ij})}{\sum_{k \in w} \exp(\beta' Y_k + \gamma Z_{ik})} \tag{3}$$

$$P_{ij|o} = \frac{\exp(\beta' Y_j + \gamma Z_{ij})}{\sum_{k \in o} \exp(\beta' Y_k + \gamma Z_{ik})} \tag{4}$$

Where *W* indicates the set of city choices within one's source province and *O* indicates the choices outside of one's source province.

4.3 Upper-level unconditional probabilities

The unconditional probabilities of the upper-level model can be estimated based on the equations:

$$P_{iw} = \frac{\exp(\alpha_w' X_i + \tau_w IV_w)}{\exp(\alpha_w' X_i + \tau_w IV_w) + \exp(\alpha_o' X_i + \tau_o IV_o)} \tag{5}$$

$$P_{io} = \frac{\exp(\alpha_o' X_i + \tau_o IV_o)}{\exp(\alpha_w' X_i + \tau_w IV_w) + \exp(\alpha_o' X_i + \tau_o IV_o)} \tag{6}$$

where *IV* are the inclusive values, also called the log-sum, which can be obtained by taking the log of the denominators of the conditional probability equations (3) and (4) for the "Within" and "Out" nest, respectively.

$$IV_w = \ln \sum_{k \in w} \exp(b'Y_k + g'Z_{ik}) \tag{7}$$

$$IV_o = \ln \sum_{k \in o} \exp(\beta' Y_k + \gamma' Z_{ik}) \tag{8}$$

Thus, the inclusive values are calculated based on the estimated coefficients from the lower-level model regressions. Subsequently, these log-sum values enter the upper-level estimation as explanatory variables. The inclusive values (*IV*) reflect the expected utility derived from migrating "Within" versus migrating "Out" of one's home province. Thus, the probability of choosing to migrate "within" vs. "Out" in the top-level model depends on the expected utility that the migrant receives from the chosen nest. The coefficient on the log-sum term τ, also referred to as "dissimilarity parameter", reflects the degree of independence for the unobserved attributes of the alternatives within the nest. A low value of τ can be viewed as high correlations between the alternatives within a nest ($\rho = 1 - \tau$), thus a higher substitution among the alternatives in a nest. On the other hand, a high value of τ indicates that the alternatives in a nest are highly independent of one another; when τ approaches 1, the nested logit collapses to the standard logit model. A nested structure is not warranted.

Note that the empirical models in this analysis, equations (3)-(6), are not identical to the specifications in the standard sequential estimation procedure because the parameters in the deterministic portion of utility function are not divided by τ. Since $1/\tau$ serves as a scale factor, the empirical equations employed here are not scaled. For this reason, it is called the non-normalized form of nested logit model. There are two alternative forms of specifications for the nested logit model: the non-normalized form developed by Ben-Akiva (1973) and the utility maximizing form developed by McFadden (1978, 1980). Both forms give consistent estimates; however, they can produce different results depending on the nesting structure and lead to a drastically different model interpretation and prediction. Koppelman and Wen (1998) compared these two forms of models side by side and concluded that the non-normalized form is consistent with the utility maximization if the log-sum parameters in the top-level branches are restricted to equality, that is, $\tau_w = \tau_o$ in this analysis. Following Koppelman and Wen (1998), we estimate the upper-level model by imposing the constraint that the log-sum coefficients are equal for the "within" and "out" nests.

4.4 Empirical results

4.3.1 Conditional logit estimation

First, we run conditional logistic regression assuming that the IIA condition is satisfied, thus the error terms of the alternatives are assumed to be independent and a standard logit model is adequate. The data need to be structured in a way that each individual in the data set faces 15 alternative destinations from which to choose. Thus, there will be 15 rows in the data array for each individual. Although our data contain both individual and choice-specific characteristics, the individual characteristics, such as age, gender and education, do not have variation within each individual, thus they cannot directly enter the conditional logit model. We can still include individual attributes in the estimation by constructing a vector of alternative dummies, in this case, one for each destination city, and by interacting the individual attributes with the set of city dummies, the conditional logit will provide the effects of personal attributes on the propensity to migrate to a particular city destination. However, when the set of alternatives is large, the parameters estimated in the model are many and the regression results can become hard to interpret.

Table (4) presents the conditional logit regression results based on working-age migrants from 16 to 64 years of age. Various city attributes are taken from the City Statistical Yearbook of China. We use the average city characteristics lagged 1-5 years prior to 2008 or 2009 since migrants supposedly made their decisions based on the previous economic conditions at the destination city. We also fit the conditional logit model for a subsample of household heads since we observe that the majority of migrant couples moved to the same city and made similar income, hence their choices cannot be viewed as separate decisions.

In the baseline specification of column (1) of Table (4) we include a set of covariates that are commonly used in the migration literature: the population size, GDP per capita, employment growth rate, distance between origin and destination location, distance squared term, and the wage differentials between rural origin and city destinations. The first three variables are major indicators of a city's economic condition and employment opportunities. As deterrence to migration, distance captures the costs associated with the moving, and its squared term is to capture the existence of any non-linear effect of distance on the probability of migrating to a particular destination. Undoubtedly, the "pull-effect" exerted by wage differential is one of the most robust findings in the migration studies. One advantage of using RUMiC data is that it provides the actual income a migrant *would have earned* had he/she stayed in their rural village. However, we can only observe the wage of a migrant at the chosen destination, the wage a migrant could have earned at alternative destinations have to be estimated. Thus, we estimate the wage for the set of alternatives (other than the one that has been chosen) by fitting a Mincer-type income equation. Specifically, the actual income of migrants is regressed on a set of individual characteristics including age, age squared, gender, education level and marital status, by city and by year, respectively. The estimated coefficients on these demographic variables are then used to generate the predicted wage a migrant could have earned for other city alternatives. Finally, wage differential is defined as the log difference of the estimated wage at each city alternative and the self-reported income at one's rural hometown prior to migration.

The results shown in Table (4) are largely consistent with many studies using data aggregated at the state/province level. All of the estimated coefficients have expected signs and are statistically significant at the 1% level. Overall, cities with a larger population size, higher per capita GDP, faster employment growth, larger wage differentials are more attractive to rural migrants. As expected, a city is less likely to attract far away migrants, as indicated by the coefficient of the distance between sending and receiving areas. Moreover, the estimated coefficient on the distance squared term is positively significant, suggesting that the effect of distance takes a U shape, that is, the deterrent effect of distance diminishes as distance increases. This makes sense because if a rural migrant from Sichuan has travelled over 1200 miles to get to the Guangdong region, he probably cares less about travelling an extra 40 miles from the city of Guangzhou to Dongguan.

In column (2) of Table (4), we further control for several additional city-specific attributes that are relevant in the context of rural-urban migration, such as the tertiary-to-primary industry employment ratio in urban area (as proxies for industry structure); the concentration of human capital (measured by the share of college graduates in the city's population); rent (measured by the logarithm of monthly expenditure on housing related expenses as a proxy for cost or standard of living); and a set of regional dummies. The 15 cities are grouped into three regions. Region 1 stands for the seven cities in the interior region, mainly provincial-level/capital cities: Zhengzhou, Hefei, Wuhan and Chengdu, Chongqing, Bengbu, and Luoyang; Region 2 is an indicator for the three cities in the Pearl River Delta: Guangzhou, Dongguan and Shenzhen; and Region 3 indicates the five cities in the Yangtze River Delta: Shanghai, Nanjing, Wuxi, Hangzhou and Ningbo. The group of interior cities (Region 1) is omitted in the regression as the reference group.

The results shown in column (2) of table (4) indicate that the coefficient on human capital is negatively significant, while the coefficient on rent is positively significant. One interpretation can be that rent reflects the amenities and the quality of life that a city offers. For example, rent is substantially higher in a city experiencing fast growth, such as Shanghai, the exact region where employment opportunities abound and attracts people from all walks of life. The estimated coefficients for the Pearl river delta and Yangtze river delta are both positive and highly significant, indicating that these two regions are much more attractive destinations for rural migrants compared with the interior region.

Also worthy of noting is that the estimated coefficient on the intra-provincial dummy is large and significant at the 1% level, suggesting there is a strong "border-effect" as migrants prefers to stay within the border of their home provinces, i.e. a city is more attractive to migrants within the same province.

To examine why there exists a significant border-effect, we explore the role of culture, institutional barrier for settlement, and dialect. First, we construct a dummy variable indicating whether a migrant has moved within a cultural district. Second, we utilize the "settlement threshold index" constructed by Wu, Zhang and Chen (2010), which reflects the degree of difficulty to permanently settle in a particular city. The higher the index is, the bigger the barrier for migrants to settle and enjoy the local public goods such as primary education and medical care. Third, we utilize the language diversity, so called "ethno-linguistic fractionalization index", to measure the impact of dialect on the propensity to migrate. As Xu, Liu and Xiao (2015) pointed out, dialect constitutes a barrier for communication, thwarts trust and promotes psychological distance, thereby discouraging migration.

As shown in column (3) of Table (4), the within-culture dummy is not significant when intra-provincial migration is controlled for. We also find that higher "settlement threshold index" is positively related to the probability of migration (consistent with Xia and Lu, 2015), and the coefficient on diversity of dialect is negatively significant (consistent with Xu et al., 2015). More importantly, the border effect becomes less pronounced as indicated by a substantial decrease in its magnitude, even though it remains statistically significant at the 1% level. This indicates that the border effect can be partly attributed to the institutional and language barriers, but they do not constitute the major factors driving the border effect we are currently observing. It is worth mentioning that the coefficient on human capital becomes insignificant after controlling for these additional covariates, suggesting a high degree of correlation between a city's concentration of human capital and its cultural, institutional, and language environments.

As shown in column (4)-(6) of Table (4), regression results based on the subsample of household heads are qualitatively similar to that of all working-age migrants, except that the effect of wage differentials becomes much larger and remains significant at the 1% level. One plausible explanation of this substantial increase in the effect of wage differential is that wage consideration is more of a priority for household heads than other family members such as spouses or children.

4.3.2 Hausman's test for IIA assumption

It is well known that conditional logit model requires IIA to be valid. To test if such a requirement is met, we perform Hausman's specification test developed by McFadden et al. (1978). The intuition behind the test is straightforward. Under IIA, the ratio of probabilities for any two alternatives is the same whether or not other alternatives are available. If IIA holds in reality, the estimated parameters obtained on a subset of alternatives will not be significantly different from those obtained on the full set of alternatives. Note that the null hypothesis in the Hausman's test is that there is no systematic difference in the parameters on the subset and those on the full set of alternatives because the alternatives are essentially independent. If the null hypothesis is rejected, there is evidence indicating that the IIA property, which is central to a standard logit model, is violated and hence a standard logit is not adequate for model specification.

However, there exists a large number ways to test IIA. For a set of 15 alternatives, if we are to exclude one alternative at a time, there are 15 ways to do so. If we exclude two alternatives at a time, there are 105 combinations, so on and so forth. To provide a general test for IIA property, we perform the Hauman's test by excluding the set of alternatives by region from the entire set of alternatives. All of the three test results indicate that IIA condition is not met. The chi-square statistics are 221, 415, 1125 for region 1, 2 and 3, respectively, with all of the p values equal to 0,000, suggesting that the alternatives are not irrelevant from one another. Thus, a nested logit structure that allows for correlations among the alternatives is warranted.

4.3.3 Two-level nested logit estimation

Following the methods specified in the equation (3) to (6), we estimate the two-level nested logit model sequentially starting from the bottom-level. The upper-level unconditional probabilities concerning the choice of staying "within" one's origin province versus moving "out" are assumed to depend on a set of individual characteristics such as age, gender, educational attainment, as well as the economic condition of one's rural hometown such as village income.

Suppose that the choice of destination at the lower-level model is influenced by the set of location-specific attributes that we have utilized in the previous conditional logit regressions. Importantly, we expect the dissimilarity coefficient (τ) on the log-sums to be positive, since higher expected utility derived from the "within" nest should increase the probability of staying within one's home province, and vice versa. For the nested logit model to be consistent with utility maximization theory, τ needs to be in the range of (0, 1). Otherwise, the nested logit model is not appropriate and does not provide consistent estimates.

Table (5) presents the maximum likelihood estimates (MLEs) of the nested logit model for the samples of working-age migrants and household heads, respectively. The estimation on the conditional probabilities in the lower-level model, conditional on the decision of "within" versus "out", generated qualitatively similar results compared to the standard logit model. All city-specific attributes are statistically significant, the majority of which have noticeably larger coefficients, with the exception that the absolute magnitude of the estimated coefficient on distance becomes much smaller, suggesting that the deterring effect of distance becomes less important given that a migrant has decided on the top-level choices. Estimating the conditional probabilities also substantially improves the fit of the model. Although not directly comparable, the log likelihood for the subsample of household heads increases from -11,767 in column (4) of Table (4) to -7,793 in column (2) of table (5), a gain of 3,974 units.

The results on the top-level model shed lights on the importance of personal attributes in influencing whether a migrant chooses to move "within" versus "out" of his/her source province. First, it's important to notice that the dissimilarity coefficient τ on the log-sum is contained in the interval (0, 1) and is statistically significant at the 1% level. This indicates that the expected utility obtained from each nest is an important determinant in the top-level choices. The correlation among the unobserved components of utility for alternatives in each nest is estimated to be 0.43 for the entire sample (recall that we have constrained the τ to be the same for both nests), though it seems to be much lower for the subsample of household heads.

Furthermore, the nested structure allows us to directly examine the statistically difference in personal attributes of the migrants who chose to move "Within" versus "Out". The migrants that migrated "Out" serve as the base group in the regression. Thus, the estimated coefficients reflect the contribution of each personal attribute on the propensity to migrate within one's home province. Consistent with the previous analysis in this paper, the results clearly show that being older, more educated, female, single, or from poorer rural areas, increases the probability of choosing a city within one's province of origin. Besides, these results are substantially different from a simple logit model using a dichotomous dependent variable indicating whether a migrant has moved within his/her home province (not presented, but available upon request).

5. Conclusions

The last several decades have witnessed large-scale rural-urban migrations in China. Concurrently, research on this topic has accelerated; however, analyses from a receiving area perspective based on micro data have been scant. This paper aims to fill this gap by employing a nested logit approach to evaluate the role of personal characteristics as well as location-specific attributes in the rural to urban migration in China.

Our results suggest that migrants from other provinces tend to be younger, more likely to be male. The receiving areas attract majority of migrants from their home province. While no similar results have been provided by the existing migration literature in China, several studies that examined the differences between circular and permanent migrants seem to provide some insight into our findings. For example, Cai and Wang (2008) and Yang (2000) find that permanent migrants tend to move within home provinces. Furthermore, Hu, Xu and Chen (2011) find that permanent migrants tend to be older, female, more educated, and more likely to become entrepreneurs, which mirrors the findings in this paper regarding the characteristics of intra-provincial migrants compared to their inter-provincial counterparts. Viewing these findings together, it may suggest that intra-provincial migrants are more likely to permanently settle in cities within their home provinces. It is highly probable that the majority of inter-provincial migrants are temporary/circular migrants, with the intention of ultimately returning to their rural origins. If this is the case, intra-provincial migration is the driving force that will continue to fuel the urbanization process in the years to come in China.

As China moves away from its export-driven growth to focus on stimulating domestic demand, the consumption and investment potentials brought by rural migrants will play a crucial role in sustaining long term economic growth in Chinese cities. When intra-provincial migrants move to cities, they supply their labor to local economies for a longer period of time, acquire new skills, generate income, purchase goods and services, and increase the demand for housing, all of which help boost the city's economic growth. There is no denying that rural-urban migration can be a double-edged sword. As the rural-urban migration continues, local governments should invest in their capacity to accommodate these large population flows and provide migrants with adequate levels of public services including affordable housing, health care and education. The facilitation of migration flows will lead to more agglomeration of economic activities, add to the stock of human capital and stimulate spending in cities. Meanwhile, local authority should also try their best to increase the information available to migrant job seekers, facilitate and guide permanent migrants to better adapt to the urban society.

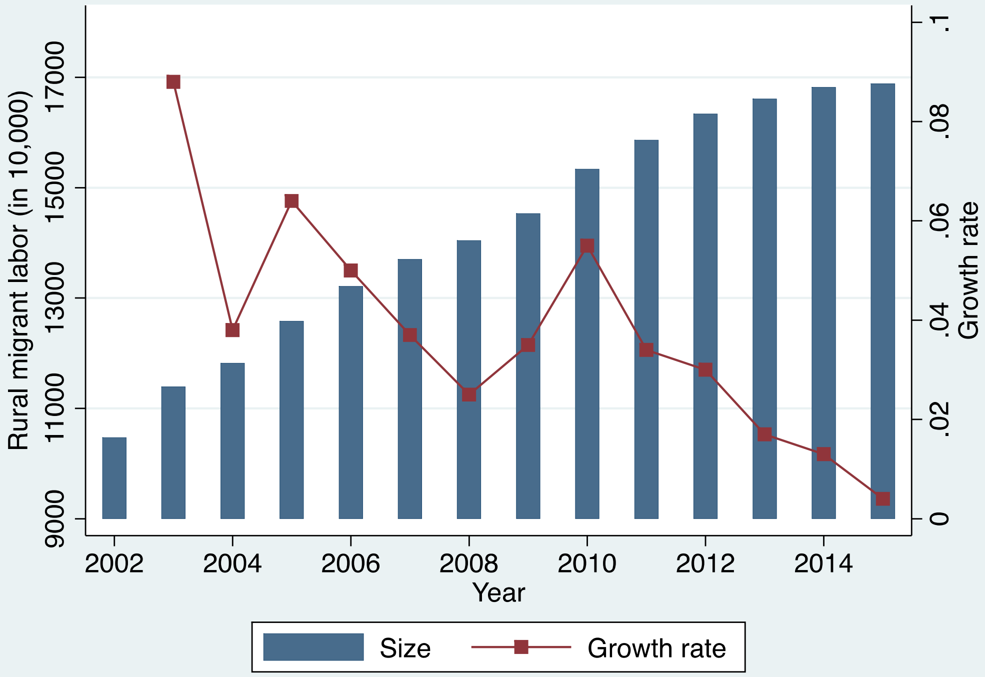

Figure 1: The size of rural migrant labor 2002-2015 and its growth rate.

Data sources: The National Bureau of Statistics (NBS).

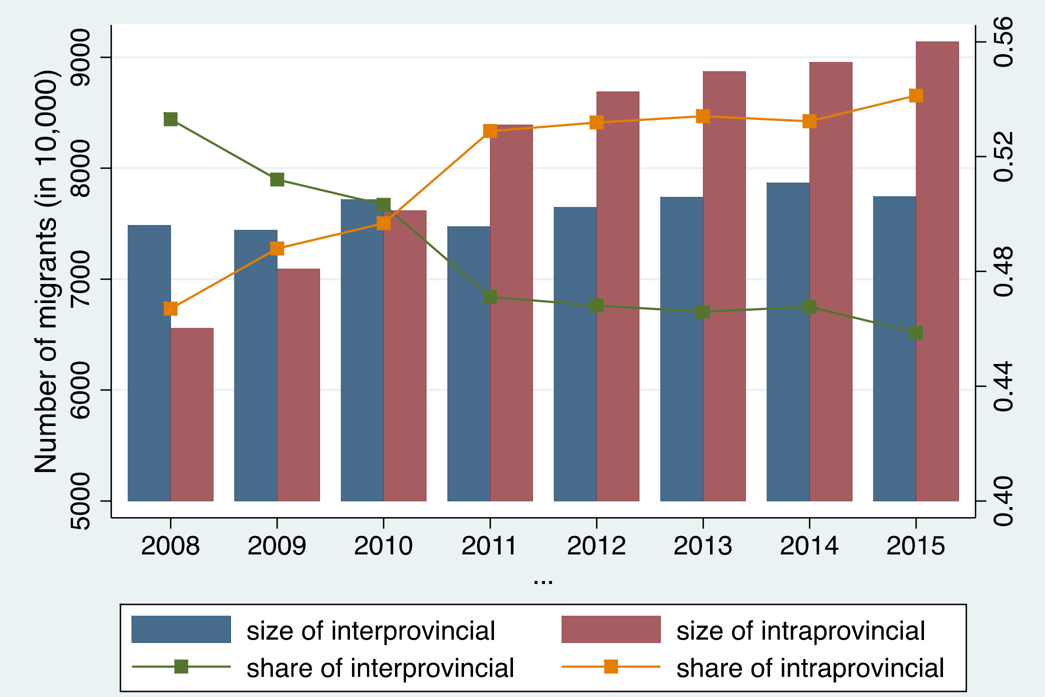

Figure 2: Size & ratio of inter- vs. intra-provincial migrants

Data sources: Survey reports on rural workers by the NBS.

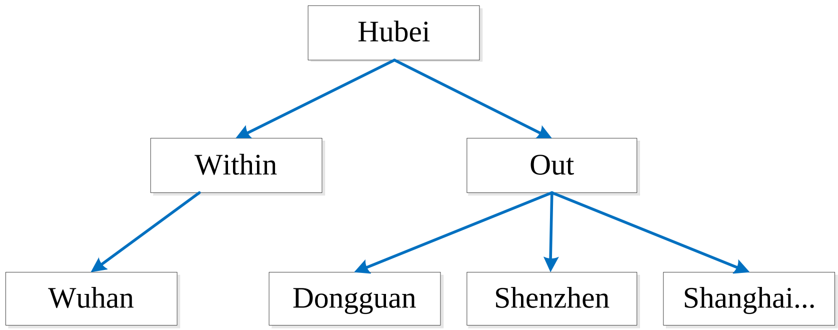

Figure 3: Two-level Nested Structure for Migrants from Hubei Province

Table 1 Individual characteristics of migrants aged 16-64

Variable	Obs	Mean	Std. Dev.	Min	Max
Age	12322	31	10.39	16	64
Young (16-35)	12322	66%	0.473	0	1
Old (35-64)	12322	34%	0.473	0	1
Male (Dummy)	12322	57%	0.495	0	1
Education (Dummy)	11817	9.2	2.508	1	20
No school	12322	2%	0.154	0	1
Primary	12322	13%	0.333	0	1
Junior High	12322	46%	0.498	0	1
Senior high	12322	21%	0.407	0	1
Vocational	12322	11%	0.312	0	1
Some college and above	12322	7%	0.259	0	1
Married (Dummy)	12322	61%	0.487	0	1
Number of children	10482	1.03	0.910	0	9
Monthly income (Yuan)	11636	1615	1629.34	0	76000
Wage_earner	9034	1420	775.410	0	10000
Self_employed	2602	2292	3032.67	0	76000
Self_employed (Dummy)	11931	22%	0.413	0	1
Village income	11707	677	503.50	0	10000
Intra-provincial (Dummy)	12322	55%	0.497	0	1
Years of migration	11819	8	6.515	0	59
Cities ever migrated	12016	2	1.937	0	50
Health: *Excellent*	12322	38%	0.486	0	1
Good	12322	46%	0.498	0	1
Average	12322	14%	0.349	0	1
Poor	12322	1%	0.120	0	1
Very poor	12322	0%	0.039	0	1

Data sources: The RUMiC Surveys 2008 and 2009 waves.

Table 2 Source of migrants for 15 cities

	Sichuan	Henan	Guangdong	Jiangsu	Zhejiang	Anhui	Hubei	Chongqing	Total
Chongqing (Chongqing)	15%	0%	0%	0%	0%	0%	0%	**84%**	99.6%
Chengdu (Sichuan)	**62%**	0%	0%	0%	0%	0%	1%	4%	66.9%
Guangzhou (Guangdong)	2%	3%	**68%**	0%	0%	0%	2%	0%	75.8%
Dongguan (Guangdong)	4%	4%	**14%**	0%	0%	0%	9%	3%	34.8%
Shenzhen (Guangdong)	5%	3%	**16%**	0%	1%	0%	5%	0%	31.1%
Shanghai (Shanghai)	3%	4%	0%	26%	15%	13%	4%	3%	68.7%
Nanjing (Jiangsu)	1%	2%	0%	**35%**	3%	10%	1%	0%	51.1%
Wuxi (Jiangsu)	1%	0%	0%	**27%**	1%	2%	0%	1%	32.1%
Hangzhou (Zhejiang)	1%	5%	0%	5%	**54%**	8%	3%	2%	79.7%
Ningbo (Zhejiang)	2%	2%	0%	3%	**17%**	6%	2%	1%	33.0%
Zhengzhou (Henan)	0%	**45%**	0%	0%	1%	2%	0%	0%	48.8%
Luoyang (Henan)	0%	**25%**	0%	0%	4%	0%	1%	0%	29.7%
Hefei (Anhui)	2%	1%	0%	1%	2%	**37%**	1%	0%	43.5%
Bengbu (Anhui)	1%	2%	1%	2%	1%	**20%**	0%	0%	26.0%
Wuhan (Hubei)	1%	3%	1%	0%	1%	0%	**71%**	2%	79.2%
Total	100%	100%	100%	100%	100%	100%	100%	100%	

Data sources: The RUMiC Surveys 2008 and 2009 waves.

Table 3 Comparison of inter-provincial and intra-provincial migrants aged 16-64

	RUMiC (n=12,322)			Census 1% (n=81,198)		
Variables	Inter-provincial	Intra-provincial	Differences in means	Inter-provincial	Intra-provincial	Differences in means
Age (Year)	31.07	31.53	-0.46**	29.64	31.48	-1.84***
Male (Dummy)	0.59	0.55	0.04***	0.56	0.57	-0.01**
Education (Year)	9.08	9.21	-0.12***	8.89	8.92	-0.04**
Married (Dummy)	0.61	0.61	0.00	0.62	0.66	-0.04***
Number of children	1.06	1.02	0.04**			
Having brought children to city (Dummy)	0.11	0.16	-0.04***			
Monthly income (Yuan)	1743	1509	233***	975	900	74***
Healthy (1 if excellent or good; 0 otherwise)	0.42	0.35	0.07***	1	0.99	0.00**
Years since first migration	7.53	8.24	-0.70***			
Cities ever migrated	2.02	1.8	0.22***			
Risk averse	4.13	4.19	-0.06			
Means to get current job						
Assigned by government (Dummy)	0	0.01	-0.00***			
Agency of government or community (Dummy)	0.08	0.06	0.01***			
Applied by self (Dummy)	0.21	0.2	0.01			
Introduced by family, relatives or friends (Dummy)	0.57	0.6	-0.03***			
Employer recruitment (Dummy)	0.1	0.08	0.03***			
Unemployment insurance (Dummy)	0.12	0.11	0.01**	0.08	0.06	0.02***
Pension (Dummy)	0.21	0.16	0.05***	0.15	0.12	0.04***
Injury insurance (Dummy)	0.19	0.13	0.06***	0.19	0.14	0.04***
Contracted worker (dummy)	0.44	0.3	0.14***	0.29	0.17	0.12***
Temporary worker (dummy)	0.2	0.26	-0.06***	0.21	0.24	-0.03***
Family business (dummy)	0.03	0.05	-0.02***	0.02	0.03	-0.02***
Self-employed (Dummy)	0.19	0.25	-0.06***	0.12	0.21	-0.09***

Data sources: The RUMiC surveys 2008 and 2009 waves; 2005 1% Census.

Table 4 Conditional logit estimates for migrants and household heads aged 16-64

	(1)	(2)	(3)	(4)	(5)	(6)
	All migrants			Household heads		
Population	0.395***	0.382***	0.680***	0.332***	0.429***	0.624***
	(0.018)	(0.022)	(0.060)	(0.021)	(0.026)	(0.074)
Per capita GDP	0.879***	0.077**	0.598***	0.740***	0.124***	0.504***
	(0.028)	(0.037)	(0.108)	(0.034)	(0.046)	(0.132)
Emp growth	0.393***	0.886***	0.956***	0.016	0.846***	0.868***
	(0.049)	(0.062)	(0.086)	(0.062)	(0.075)	(0.103)
Distance	-0.005***	-0.004***	-0.005***	-0.005***	-0.004***	-0.005***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Distance sq	0.016***	0.012***	0.013***	0.016***	0.012***	0.013***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)
Wage Diff.	0.075***	0.047***	0.056***	1.649***	1.710***	1.748***
	(0.015)	(0.016)	(0.017)	(0.047)	(0.051)	(0.057)
Emp ratio		0.935***	-1.457***		1.218***	-1.620***
		(0.063)	(0.307)		(0.075)	(0.381)
Human capital		-0.121***	0.044		-0.142***	0.052
		(0.006)	(0.029)		(0.008)	(0.035)
Rent		0.397***	0.297***		0.266***	0.094
		(0.051)	(0.101)		(0.061)	(0.117)
2.Region		1.700***	0.769***		1.091***	0.118
		(0.076)	(0.132)		(0.094)	(0.164)
3.Region		0.834***	0.055		0.373***	-0.485***
		(0.048)	(0.134)		(0.059)	(0.167)
Intra-provincial		1.271***	1.001***		1.358***	1.060***
		(0.053)	(0.060)		(0.065)	(0.075)
Within-culture			0.071			0.030
			(0.070)			(0.089)
Hukou threshold			0.324***			0.510***
			(0.108)			(0.132)
Dialect			-1.877***			-2.273***
			(0.193)			(0.235)
pseudo R^2	0.354	0.389	0.443	0.397	0.430	0.487
Log lik.	-17526.829	-16581.083	-11988.548	-11767.319	-11124.002	-7949.743
N	150255	150255	103956	108180	108180	74808

Notes: City attributes (Population, per capita GDP, Emp ratio, and human capital) are averages over five-year period based on China's urban statistical yearbooks, i.e., 2003-2008 for the migrants in 2008 RUMiC survey and 2004-2009 for the migrants in 2009 RUMiC survey. Population is the total urban population (in 10,000). Per capita GDP is the GDP divided by total population. Emp growth is the employment growth rate over the 5-year period ending in the RUMiC survey year. Distance is the railway distance between the capital city of source province and the current working city (in Kilometer). Wage differential is the differential between a migrant's income in the rural area of origin and that in the working city. Emp ratio is the tertiary-to-primary industry employment ratio in urban area (as proxies for industry structure). Human capital is the share of college graduates in city population. Rent is the logarithm of migrant households' monthly housing related expenditures at the working city. Region 1 stands for the seven cities in the interior region: Zhengzhou, Hefei, Wuhan and Chengdu, Chongqing, Bengbu, and Luoyang; Region 2 is an indicator for the three cities in the Pearl River Delta: Guangzhou, Dongguan and Shenzhen; and Region3 indicate the five cities in the Yangtze River Delta: Shanghai, Nanjing, Wuxi, Hangzhou and Ningbo. The group of interior cities (Region1) is omitted in the regression as the reference group. Intra-provincial is a dummy variable indicating a migrant has moved within home province; within-culture is a dummy variable indicating a migrant has moved within cultural district. *Hukou* threshold is the "settlement threshold index" constructed by Wu, Zhang and Chen (2010). Dialect is the "ethno-linguistic fractionalization index" used in Xu, Liu and Xiao (2015).

Robust standard errors are reported in parentheses. ***, **, and * denote significance at the .01, .05, and .10 levels, respectively.

Data sources: The RUMiC surveys 2008 and 2009 waves; China's urban statistical yearbook.

Table 5 Maximum likelihood estimates for two-level nested logit model

	all migrants aged 16-64	household head
	Upper-level probabilities : "within" versus "out" choice	
Log-sum	0.573***	0.825***
	(0.030)	(0.033)
Age (Year)	0.122***	0.145***
	(0.012)	(0.015)
Age squared	-0.001***	-0.002***
	(0.000)	(0.000)
Education (Year)	0.076***	0.078***
	(0.010)	(0.013)
Male (Dummy)	-0.062	-0.136**
	(0.048)	(0.067)
Married (Dummy)	-0.365***	-0.374***
	(0.069)	(0.085)
Village income (Yuan)	-0.256***	-0.258***
	(0.030)	(0.038)
	Lower-level probabilities: city destination choice	
Population	0.603***	0.567***
	(0.021)	(0.026)
Per Capita GDP	0.901***	0.777***
	(0.031)	(0.038)
Emp Growth	0.348***	0.108
	(0.059)	(0.072)
Wage Differential	0.103***	1.492***
	(0.021)	(0.056)
Distance	-0.003***	-0.003***
	(0.000)	(0.000)
Distance Squared	0.007***	0.007***
	(0.000)	(0.001)
Log likelihood	-11458.046	-7793.632
N	62061	44689

Notes: Log-sum refers to the inclusive values. Village income is the average monthly income of one's rural hometown. Other covariates are the same as previously defined in Table 4.

Robust standard errors are reported in parentheses. ***, **, and * denote significance at the .01, .05, and .10 levels, respectively.

Data sources: RUMiC surveys 2008 and 2009 waves; China's urban statistical yearbook.

References

Ao, X., Jiang, D., & Zhao Z. (2016). The impact of rural–urban migration on the health of the left-behind parents. *China Economic Review*, 37, 126-139.

Au, C. C., & Henderson, J. V. (2006). How migration restrictions limit agglomeration and productivity in china. *Journal of Development Economics*, 80, 350–388.

Bao, S., Bodvarsson, O. B., Hou, J. W., & Zhao, Y. (2007). Interprovincial migration in China: the effects of investment and migrant networks. *Technical Report IZA Discussion Papers*.

Ben-Akiva, M. E. (1974). Structure of passenger travel demand models. *Transportation Research Record*.

Cai, F., & Wang, M. (2010). Growth and structural changes in employment in transition China. *Journal of Comparative Economics*, 38(1), 71-81.

Cai, H., & Wang, J. (2008). Factors influencing the migration intentions of rural workers in the pearl river delta. *Social Sciences in China*, 29, 157–171.

Chan, K. W. (1994). Urbanization and rural-urban migration in China since 1982: a new baseline. *Modern China*, 20(3), 243-281.

Chan, K. W., & Buckingham, W. (2008). Is China abolishing the hukou system?. *The China Quarterly*, 582-606.

Chan, K. W., & Zhang, L. (1999). The hukou system and rural-urban migration in China: Processes and changes. *The China Quarterly*, 160, 818-855.

Chang, G. H., & Brada, J. C. (2006). The paradox of china's growing under- urbanization. *Economic Systems*, 30, 24–40.

Chen, Y. P. (2012). Land use rights, market transitions, and labour policy change in China (1980–84). *Economics of Transition*, 20(4), 705-743.

Cheng, M., Shi, Q., & Ma, L. (2006). An explanation of the motivation and obstacles affecting rural-urban migration: The case of china. *Frontiers of Economics in China*, 1, 576–596.

de Brauw, A., & Mu, R. (2011). Migration and the overweight and underweight status of children in rural china. *Food Policy*, 36, 88 – 100.

Garnaut, R., & Song, L. (2013). The turning point in China's economic development (p. 390). *ANU Press.*

Han (2007). Government should adopt policies favouring rural migrants. *China Economist*, 6, 30–41.

Hu, F., Xu, Z., & Chen, Y. (2011). Circular migration, or permanent stay? evidence from china's rural–urban migration. *China Economic Review*, 22, 64–74.

Koppelman, F. S., & Wen, C.-H. (1998). Alternative nested logit models: structure, properties and estimation. *Transportation Research Part B: Methodological*, 32, 289–298.

Lee, M. H. (2011). Migration and children's welfare in China: The schooling and health of children left behind. *The Journal of Developing Areas*, 44(2), 165-182.

Liang, Z. (2001). The age of migration in China. *Population and development review*, 27(3), 499-524.

Liang, Z., & Ma, Z. (2004). China's floating population: new evidence from the 2000 census. *Population and development review*, 30(3), 467-488.

Liang, Z., Li, Z., & Ma, Z. (2014). Changing patterns of the floating population in China, 2000–2010. *Population and development review*, 40(4), 695-716.

Lin, J. Y., Wang, G., & Zhao, Y. (2004). Regional inequality and labor transfers in china. *Economic Development and Cultural Change*, 52, 587–603.

Ma, Z., & Liaw, K. L. (1997). Explaining hierarchical and interprovincial migrations of Chinese young adults by personal factors and place attributes: a nested logit analysis. *Mathematical Population Studies*, *6*(3), 217-239.

Ma, Z., Liaw, K. L., & Zeng, Y. (1997). Migrations in the urban-rural hierarchy of China: Insights from the microdata of the 1987 National Survey. *Environment and Planning A*, 29(4), 707-730.

Manski, C. F., McFadden, D. et al. (1981). Structural analysis of discrete data with econometric applications. *Mit Press Cambridge*, MA.

McFadden, D. (1980). Econometric models for probabilistic choice among products. *Journal of Business*, 53, S13–S29.

McFadden, D. et al. (1978). Modelling the choice of residential location. *Institute of Transportation Studies*, University of California.

Meng, X. (2000). Labour market reform in China. *Cambridge University Press*.

Murphy, R., Zhou, M., & Tao, R. (2015). Parents' Migration and Children's Subjective Well-being and Health: Evidence from Rural China. *Population, Space and Place*.

Poncet, S. (2006). Provincial migration dynamics in china: Borders, costs and economic motivations. *Regional Science and Urban Economics*, 36, 385–398.

Seeborg, M. C., Jin, Z., & Zhu, Y. (2000). The new rural-urban labor mobility in china:

Causes and implications. *Journal of Socio-Economics*, 29, 39 – 56.

Song, Y., & Sun, W. (2015). Health Consequences of Rural-to-Urban Migration: Evidence from Panel Data in China. *Health economics*.

Taylor, G. (2011). China's floating migrants: Updates from the 2005 1% population sample survey. *London School of Economics Migration Studies Center Working Paper*, (2011/07).

Wu K. Y., Zhang L., & Chen X. (2010). Huji Gaige Jincheng De Zhangai: Jiyu Chengshi Luohu Menkan De Fenxi (in Chinese). *Chinese Journal of Population Science*, 01, 66-74+112

Wu, H. X., & Zhou, L. (1996). Rural-to-Urban Migration in China. Asian-Pacific *Economic Literature*, 10(2), 54-67.

Xia Y. R., & Lu M. (2015). Chengshi Jian De "Mengmu Sanqian"—Gonggong Fuwu Yingxiang Laodongli Liuxiang De Jingyan Yanjiu (in Chinese), *Management World*, 10, 78-90.

Xu X. X., Liu M. Y., & Xiao Z. K. (2015). Fangyan Yu Jingji Zengzhang (in Chinese), *China Journal of Economics*, 02, 1-32.

Xu, H., & Xie, Y. (2015). The causal effects of rural-to-urban migration on children's well-being in China. *European sociological review*, jcv009.

Yang, H. (2000). A comparative analysis of china's permanent and temporary mi- gration during the reform period. *International Journal of Social Economics*, 27, 173–193.

Zhao, Y. (1999a). Labor migration and earnings differences: the case of rural china. *Economic Development and Cultural Change*, 47, 767–782.

Zhao, Y. (1999b). Leaving the countryside: rural-to-urban migration decisions in china. *American Economic Review*, (pp. 281–286).

Zhu, N. (2002). The impacts of income gaps on migration decisions in china. *China Economic Review*, 13, 213–230.