The Preference for Larger Cities in China: Evidence from Rural-Urban Migrants

Chunbing Xing and Junfu Zhang[*]

August 9, 2013

Abstract

China has long aimed to restrict population growth in large cities but encourages growth in small and medium-sized cities. At the same time, various government policies favor large cities. We conjecture that larger cities in China have more urban amenities and a better quality of life. We thus predict that a typical rural-urban migrant is willing to give up some income in order to live in a larger city. We present a simple model in which rural-urban migrants choose destination cities to maximize utilities from consumption and urban amenities. Drawing data from a large-scale population survey conducted in 2005, we first estimate each migrant's expected earnings in each possible destination city using a semi-parametric method to correct for potential selection bias. We then estimate the typical migrant's preference for city population size, instrumenting population size with its lagged values to control for potential omitted-variables bias. From these estimation results, we calculate the typical migrant's willingness to pay to live in larger cities. Our results show that indeed rural-urban migrants strongly prefer cities with larger populations. We explore possible explanations for this preference and discuss the implications of these findings.

Keywords: City size, urban amenities, rural-urban migration, *hukou* system, China.

JEL Classification: O15, R12, R23.

1   Introduction

Cities come in different sizes. Traditional urban economic theory explains the distribution of city sizes based on various agglomeration economies and diseconomies (Henderson, 1974). However, such economic forces are not the only determinants of city sizes; political factors sometimes feature more predominantly. For example, dictators may invest disproportionately more in their capital cities for political stability concerns, producing urban giants that are hard to explain on pure economic grounds (Ades and Glaeser, 1995). Similarly, in a planned economy, policy makers—in both central and local governments—can influence city sizes through investment decisions and migration controls (Au and Henderson, 2006a, 2006b).

China provides an interesting case for studying the effect of government policies on city sizes. During 1949-1992, China was officially a planned economy, where central and local governments always intended to manage city growth through planning and regulations. As is well known, China has for decades had a residence registration system, which controls internal population migration and makes it particularly difficult for rural people to move into cities. Meanwhile, as a result of government planning, major industries are dispersed over different regions. Consequently, industrial clusters are relatively small and fail to take full advantage of localization economies (Lu and Tao, 2009). In addition, China has long encouraged the growth of small and medium-sized cities, and contained the growth of population in large cities (Henderson, 2005).

Along with its remarkable economic growth, China has experienced a rapid urbanization in the past three decades. While only 18 percent of the population lived in urban China in 1978, over 50 percent reside in cities today. This is mainly a result of relaxing the control of internal migration and accommodating labor mobility required by fast-growing urban sectors. During this period, the Chinese government has continued to curb population growth in large cities and direct migrants to smaller cities. At the same time, economic development policies favor large cities. For example, larger cities receive more investment, are granted more political power, and enjoy more freedom in managing local development. As a result, the quality of life in larger cities tends to be higher.

In this paper, we empirically show that larger cities in China are more attractive, as evidenced by the revealed preferences of rural-urban migrants. To guide the empirical analysis, we present a simple model in which rural-urban migrants choose destination cities by trading off expected income (and thus consumption) for urban amenities. Drawing data from a large population survey, we first estimate each migrant's expected earnings in different cities using a semi-parametric method to correct for potential selection bias. Based on actual migration choices, we next estimate the typical migrant's willingness to pay for living in different cities. This willingness to pay is then regressed on city population size to quantify the preference for larger cities. To address potential omitted-variables bias in the city-level regression, we instrument city population with its lagged values. Our results

show that rural-urban migrants are willing to give up a substantial amount of income in order to live and work in larger cities. Observed city characteristics explain little of this willingness to pay. We explore deeper reasons why migrants prefer larger cities and discuss policy implications of these findings.

The main contribution of this study is to demonstrate the consequence of some policy distortions in the urbanization process of China, which helps us better understand the growth path of this major developing country. On the methodology side, we treat city size as a nonmarket city amenity and implement a new method to assess the value of this amenity. Traditionally, the value of city amenities is measured within the Rosen-Roback framework, which assumes zero moving costs for economic agents (e.g., Roback, 1982; Blomquist et al., 1988; Albouy, 2012). This approach has limited application for a country like China, where migration costs are prominent. In a seminal paper, Bayer et al. (2009) propose an alternative method to evaluate nonmarket amenities. They estimate a discrete choice model of migration to measure the value of clean air in U.S. cities, explicitly incorporating moving costs into the model. Timmins (2007) uses this method to quantify the value of climate amenities in Brazil. We believe that this discrete choice approach is particularly useful for studying urban amenities and related issues in China, and this paper serves as an illustration. On the data source side, we make use of a large survey database created by the National Bureau of Statistics of China, which allows us to examine detailed migration choices of a very large number of rural-urban migrants. This helps us better understand internal migration patterns in China.

In the next section, we briefly introduce the institutional context in China. We then present a simple model to provide a structural framework for empirical estimation and interpretation of results. After a brief introduction of data sources, we present our estimation results. Finally, we conclude with a few remarks.

2 Institutional Background in China

In the 1950s, China established a residence registration system. Each household is required to register their residence at a local government agency, which essentially grants each family member a residence permit (*hukou*) in the place. A newborn's *hukou* status is generally inherited from the parents, almost always from the mother in early years. In cities, this residence permit not only allows the person to reside in the local jurisdiction, but also grants access to local public school and healthcare systems. In early years, it even came with guaranteed job opportunities, subsidized grain supply, and permits to purchase rationed goods such as bicycles, sewing machines, and family electronics. In countryside, the residence permit entitles a person to reside in the area and farm on the land owned by the local economic collective; it similarly grants access to local public school and healthcare. It requires residents to provide labor service and pay head taxes and fees for local public

works. Over time, the residence registration system evolves along with economic reforms. For example, in cities, subsidized grain supply and rationed goods do not exist anymore because planned allocation has been replaced by market transactions; in countryside, land has been distributed to and farmed by individual families. All these developments affected the rights and responsibilities associated with a *hukou* (Chan and Zhang, 1999).

Whereas the main purpose of the residence registration system is to facilitate government administration, it imposes a stringent constraint on the internal migration of population. In countryside, changing *hukou* from one village to another occurs mainly for marriage reasons. In urban areas, changes across cities may occur as a result of government-authorized job transfers. Converting a person's rural *hukou* status to an urban one rarely happens, which is possible only under some specific situations, including for example: (1) college graduates, who grew up in countryside and had a rural *hukou* before entering college, are granted an urban *hukou* if working in the urban sectors; (2) state-owned enterprises recruit workers from countryside; (3) urban governments recruit cadres from countryside; (4) demobilized military personnel, if working in the urban sectors, are granted an urban *hukou*; (5) family reunions that involve a member changing residence from rural to urban area. Except channel (1), other changes occur occasionally and often require lengthy bureaucratic procedures. This control of rural-urban migration was particularly tight during the early years, thus city population grew relatively slowly in China during the pre-reform era.

After the inception of economic reform in 1978, the fast-growing urban sector, especially in the coastal regions, increased the demand for cheaper labor from the rural sector. At the same time, reforms in rural areas through the household responsibility system greatly improved productivity in agriculture, releasing a large amount of surplus labor in countryside. As a pragmatic policy response, China started to allow some of the rural people to migrate to cities on a temporary basis, without granting them the urban *hukou*.

Starting in 1980, China officially pursued a policy that "contains the scale of large cities, reasonably develops medium-sized cities, and aggressively promotes the growth of small cities." Government policies repeatedly advocated that surplus labor in rural areas should "move away from the soil but not the village, enter the factories but not cities." Consistent with this policy, low-tech, low-skill industries were encouraged in townships and small cities all over the country to absorb rural surplus labor in nearby areas. Promoting population growth of small cities remained the guiding principle in important government policies until the 2000s.[1]

In the meantime, various government policies favored larger cities. A main reason behind this is the political hierarchy among cities in China (Fujita et al., 2004). There are four large "direct-control cities," Beijing, Shanghai, Tianjin, and Chongqing, that have the same political status as provinces and autonomous regions.[2] Then there are more than

---

[1]See, for example, the Tenth Five-Year Plan of China that was passed in 2001.

[2]Chongqing was a direct-control city in the early 1950s. It was then demoted to a prefecture-level city in Sichuan province. It regained the direct-control status in 1997.

three hundred prefecture-level cities, which each administratively controls a city proper as well as its surrounding rural areas. Some of these prefecture-level cities are so large and economically significant that they are designated as the "separate-planning cities." Their economic plans are more directly controlled by the central government and their mayors generally have the same political status as vice governors of provinces within the Communist Party's hierarchy of cadres. At the lower level, there are hundreds of county-level cities, which as a jurisdiction usually have smaller urban areas and populations and control a much smaller surrounding rural area. Many of these county-level cities used to be small townships and were promoted to a city status only in recent years.

Because of this political hierarchy, larger cities are generally governed by more powerful political leaders who have the bargaining power to secure more investments, negotiate for more favorable policies from upper-level governments, and maintain a higher level of autonomy. Consequently, larger cities receive more investment, have better infrastructure, have better human capital, and are generally able to grant more favorable policies to domestic and foreign business investors. Consider one example: Most high educational institutions in China are national and state universities supported by government funds. Because universities themselves are large employers and may help improve human capital, they are often a major contributor to economic development in a city. In China, almost all major universities are located in large cities, indicating the power of large cities in obtaining government investment.

Large cities in China are designated as "leaders" in economic development. They are supposed to "lead" smaller cities, which in turn will "lead" townships and villages. Here "leaders" means that large cities will host the high-tech, knowledge-intensive industries; they will pass down the labor-intensive and more polluting industries to smaller cities or townships. They are often chosen by government agencies as the hosts of major domestic development projects and foreign direct investments.

The restriction on population growth at large cities, combined with more favorable policy treatments by central and provincial governments, implies that large cities in China have more urban amenities and a better quality of life. Consider a newcomer in urban China. When facing the choice of selecting a destination city, all else equal, she must strictly prefer a larger city. We thus expect that she is willing to give up some income in order to live in larger cities.

The massive rural-urban migration in China during recent years provides a context to test this implication. A rural-urban migrant, in principle, may choose any city as destination; she is not granted an urban *hukou* in any city. Thus we will examine whether she is indeed willing to forgo some potential earnings to choose a larger city.[3]

---

[3]Existing literature has emphasized the implication of a different feature of the urbanization process in China, namely, the overly dispersion of economic activities in urban sectors. It argues that Chinese cities are too small and that localization and urbanization economies are not fully realized (Au and Henderson, 2006a, 2006b). This implies that typical workers in a larger city will be more productive and may be paid

3 Model

We present a model of migration destination choice to provide a structural framework for empirical analysis.

Consider a group of individuals who have decided to migrate from rural to urban areas. An individual i may choose to live and work in one of J cities. If living in city j, individual i faces the following utility-maximization problem

$$\max U_{ij} = C_{ij}^{\alpha_C} H_{ij}^{\alpha_H} \exp\left[\beta_S \ln S_j + \sum_{k=1}^{K} \beta_k \ln X_{jk} + M_{ij} + \xi_j + \eta_{ij}\right] \tag{1}$$
$$s.t. \quad C_{ij} + p_j H_{ij} = I_{ij}.$$

C_{ij} is i's consumption of a tradable composite good in city j; its price is the same everywhere and normalized to 1. H_{ij} is i's consumption of a nontradable composite good (including, e.g., housing) in city j; its price in city j is p_j. S_j is the population size of city j, the key variable of interest in this study. X_{jk}, $k = 1, ..., K$, is a vector of observed characteristics of city j. M_{ij} represents a non-monetary cost of migration that is related to the distance from i's home village to city j. ξ_j captures unobserved characteristics (e.g., migrant-friendliness) of city j. η_{ij} is i's idiosyncratic component of utility, assumed to be independent of migration distance and city characteristics. And finally, I_{ij} is i's income in city j.

Given the Cobb-Douglas utility function, in any city j, i's demand for the tradable and nontradable goods will be

$$C_{ij}^* = \frac{\alpha_C I_{ij}}{\alpha_C + \alpha_H}; \quad H_{ij}^* = \frac{\alpha_H}{\alpha_C + \alpha_H} \frac{I_{ij}}{p_j}.$$

Plug these demand functions into the utility function to get the indirect utility

$$U_{ij}^* = A p_j^{-\alpha_H} I_{ij}^{\alpha} \exp\left[\beta_S \ln S_j + \sum_{k=1}^{K} \beta_k \ln X_{jk} + M_{ij} + \xi_j + \eta_{ij}\right],$$

where $A \equiv \left(\frac{\alpha_C}{\alpha_C+\alpha_H}\right)^{\alpha_C} \left(\frac{\alpha_H}{\alpha_C+\alpha_H}\right)^{\alpha_H}$ and $\alpha \equiv \alpha_C + \alpha_H$. Rescaling by $\frac{1}{A}$ and taking natural logs, we rewrite the indirect utility function as

$$V_{ij} = -\alpha_H \ln p_j + \alpha \ln I_{ij} + \beta_S \ln S_j + \sum_{k=1}^{K} \beta_k \ln X_{jk} + M_{ij} + \xi_j + \eta_{ij}. \tag{2}$$

The price of nontradable goods, p_j, is not directly observable. Following Timmins

---

more. Here we are focusing on migrant workers who do not have equal access to urban sector jobs as the local workers with an urban *hukou*. For this reason, we believe that migrant workers do not necessarily earn more even if they are more productive in larger cities.

(2007), we assume this price to be a linear function of observed city characteristics:[4]

$$\ln p_j = \lambda_s \ln S_j + \sum_{k=1}^{K} \lambda_k \ln X_{jk} + \tau_j. \tag{3}$$

Substituting into equation (2) yields

$$
\begin{aligned}
V_{ij} &= \alpha \ln I_{ij} + (\beta_S - \alpha_H \lambda_s) \ln S_j + \sum_{k=1}^{K} (\beta_k - \alpha_H \lambda_k) \ln X_{jk} + M_{ij} + (\xi_j - \alpha_H \tau_j) + \eta_{ij} \\
&= \alpha \ln I_{ij} + \beta_S^* \ln S_j + \sum_{k=1}^{K} \beta_k^* \ln X_{jk} + M_{ij} + \xi_j^* + \eta_{ij}, \tag{4}
\end{aligned}
$$

where $\beta_S^* \equiv \beta_S - \alpha_H \lambda_s$, $\beta_k^* \equiv \beta_k - \alpha_H \lambda_k$, and $\xi_j^* \equiv (\xi_j - \alpha_H \tau_j)$.

Denote WTP_i (i's marginal *willingness to pay*) as the amount of money person i is willing to give up in order to have one more unit of city population S_j. From equation (4), this willingness to pay equals the marginal rate of substitution (in absolute value) between city population and income, i.e.,

$$WTP_i = \frac{\partial V_{ij}/\partial S_j}{\partial V_{ij}/\partial I_{ij}} = \frac{\beta_S^*}{\alpha} \frac{I_{ij}}{S_j}.$$

This marginal willingness to pay is higher when i has a higher income; it is lower when i is living in a city with a larger population. Moreover, WTP_i is higher when the ratio $\frac{\beta_S^*}{\alpha}$ is higher. Alternatively, one could also measure a person's preference for larger cities using the income city-size elasticity:

$$\frac{\Delta I_{ij}/I_{ij}}{\Delta S_j/S_j} \approx \frac{\partial \ln I_{ij}}{\partial \ln S_j} = \frac{\beta_S^*}{\alpha},$$

which implies that if city population increases by one percent, a person is willing to give up $\frac{\beta_S^*}{\alpha}$ percent of her income. Either way, α and β_S^* are the key parameters needed to measure the value of a larger city population.

Individual i's income I_{ij} is not observed for every city j. Following Bayer et al. (2009) and Timmins (2007), we decompose log income into a predicted mean and an idiosyncratic error term:

$$\ln I_{ij} = \ln \hat{I}_{ij} + \varepsilon_{ij}. \tag{5}$$

We will estimate $\ln \hat{I}_{ij}$ based on individual i's characteristics and the earnings of migrants who are observed in city j, controlling for potential self-selection biases. This estimation procedure will be explained in detail in the next section.

We assume that migration cost M_{ij} varies with migration distance. A longer migration

---

[4]One could easily derive a relationship between the price of nontradable goods and urban amenities from the Rosen-Roback framework (Roback, 1982). Here we impose a linear relationship.

almost surely takes more time and efforts. And more importantly, a longer migration tends to disrupt the social-family network and puts one in an unfamiliar environment, which is likely to entail a higher psychic cost. To capture these effects, we assume that

$$M_{ij} = \pi_D \ln D_{ij} + \pi_1 d_{ij}^1 + \pi_2 d_{ij}^2, \tag{6}$$

where D_{ij} is the physical distance between i's home village and city j; $d_{ij}^1 = 1$ if city j is in a province adjacent to i's home province, and 0 otherwise; $d_{ij}^2 = 1$ if city j is in neither i's home province nor its adjacent provinces, and 0 otherwise. The two dummy variables allow for extra migration costs when one moves outside of home province.

Substitute equations (5) and (6) into (4) to get

$$V_{ij} = \alpha \ln \hat{I}_{ij} + \beta_S^* \ln S_j + \sum_{k=1}^{K} \beta_k^* \ln X_{jk} + \pi_D \ln D_{ij} + \pi_1 d_{ij}^1 + \pi_2 d_{ij}^2 + \xi_j^* + \upsilon_{ij}, \tag{7}$$

where $\upsilon_{ij} \equiv \alpha \varepsilon_{ij} + \eta_{ij}$. In principle, at this point, one could make an assumption about the distribution of υ_{ij} and estimate $(\alpha, \beta_S^*, \beta_1^*, \ldots, \beta_K^*, \pi_D, \pi_1, \pi_2)$ by maximum likelihood. However, city population S_j is likely to be correlated with many unobserved city characteristics in ξ_j^*. For example, a city with a larger population may have many migrant-friendly policies that are unobserved. If individual migration choices are influenced by these unobserved policies, the estimate of β_S^* will be biased.

The standard approach to dealing with this problem is to use a two step method: In step one, use a city fixed effect to capture the utilities derived from both observed and unobserved city characteristics. In step two, regress the city fixed effects on observed city characteristics at the city level, where one can instrument for city population size to obtain a consistent estimate of β_S^*. This is the approach we will follow here.

Let $\theta_j \equiv \beta_S^* \ln S_j + \sum_{k=1}^{K} \beta_k^* \ln X_{jk} + \xi_j^*$. We rewrite the indirect utility function in equation (7) as

$$V_{ij} = \alpha \ln \hat{I}_{ij} + \pi_D \ln D_{ij} + \pi_1 d_{ij}^1 + \pi_2 d_{ij}^2 + \theta_j + \upsilon_{ij}. \tag{8}$$

Note that everything in θ_j is fixed at the city level, so we will refer to θ_j as the city fixed effect. It represents the utility a typical migrant derives from living in city j. Properly rescaled, it can also be interpreted as the typical migrant's willingness to pay for working and living in city j.

To facilitate estimation in this step, we assume that υ_{ij} follows an i.i.d. type I extreme value distribution, giving a standard conditional logit model (McFadden, 1974, 1978). It follows that individual i chooses city j with probability

$$\Pr\left(\ln V_{ij} > \ln V_{ik} \forall k \neq j\right) = \frac{\exp\left(\alpha \ln \hat{I}_{ij} + \pi_D \ln D_{ij} + \pi_1 d_{ij}^1 + \pi_2 d_{ij}^2 + \theta_j\right)}{\sum_{s=1}^{J} \exp\left(\alpha \ln \hat{I}_{is} + \pi_D \ln D_{is} + \pi_1 d_{is}^1 + \pi_2 d_{is}^2 + \theta_s\right)}.$$

Given the assumption of independent migration decisions, the probability that every migrant

8

i is living in city j as observed in the data is given by

$$L = \prod_i \prod_{j=1}^J \left[\frac{\exp\left(\alpha \ln \hat{I}_{ij} + \pi_D \ln D_{ij} + \pi_1 d_{ij}^1 + \pi_2 d_{ij}^2 + \theta_j\right)}{\sum_{s=1}^J \exp\left(\alpha \ln \hat{I}_{is} + \pi_D \ln D_{is} + \pi_1 d_{is}^1 + \pi_2 d_{is}^2 + \theta_s\right)} \right]^{\kappa_{ij}}, \tag{9}$$

where κ_{ij} is an indicator function that equals 1 if individual i is observed in city j. We can thus estimate $\{\alpha, \pi_D, \pi_1, \pi_2, \theta_1, \ldots, \theta_J\}$ by maximizing this likelihood function. Note that if any set of parameters maximizes the likelihood function, then adding a constant to every θ_j will do the same. That is, the absolute scales of $\{\theta_1, \ldots, \theta_J\}$ are not identified. In practice, we will set $\theta_1 = 0$ (for Beijing) and interpret each of the estimated θ_j as the difference from θ_1 .

In step two, we estimate $\beta_S^*, \beta_1^*, \ldots, \beta_K^*$ from the following linear equation

$$\theta_j = \beta_S^* \ln S_j + \sum_{k=1}^K \beta_k^* \ln X_{jk} + \xi_j^*. \tag{10}$$

As mentioned above, observed city size S_j and unobserved city characteristics ξ_j^* are likely to be correlated. As a solution, we will instrument for city size.

4   Data

Our data on rural-urban migrants are drawn from the the 2005 One-Percent Population Survey of China. Since the mid-1980s, China's National Bureau of Statistics (NBS) has conducted large-scale population surveys (also known as the "mini-census") during inter-census years, typically in the 5th year after a population census. The 2005 One-Percent Population Survey is the latest of such surveys.

This survey used a long questionnaire to solicit very detailed demographic, geographic, economic, and housing information about household members. For example, we know whether a household member is working, her age, education level, monthly earnings, etc., which are crucial for estimating a person's potential earnings in different cities. We also have detailed information about a household's housing conditions such as age of the building, number of rooms, size of living area, kitchen type, whether it uses natural gas, etc. Note that although the regular population census in China has better coverage, it asks far fewer questions than this One-Percent Population Survey. The latest 2010 census does not even ask about monthly earnings. Therefore, for our purpose here, the One-Percent Population Survey is actually more useful.

Another feature of the 2005 Survey is that it was specially designed to capture population flows. It not only asks about a person's current residence, but also her *hukou* place and whether she has left the *hukou* place for more than 6 months. This information is crucial because it enables us to identify rural-urban migrants. Specifically, we classify an individual as a rural-urban migrant if this person has a rural *hukou* but currently lives and works in

a city.

For some unknown reasons, the NBS of China only agrees to release of a one-fifth random subsample of the 2005 One-Percent Population Survey data.[5] This sample contains about 2.3 million individuals, covering all 31 province-level jurisdictions. We first construct a sample of rural-urban migrants from the survey data.

A person is included in this migrant sample if he or she satisfies all of the following conditions:

(i) holds a rural *hukou* but has left the *hukou* registration place for more than 6 months;

(ii) has migrated out of rural area for employment reasons;

(iii) is currently living in an urban area;

(iv) is between 20 and 60 years old;

(v) is currently employed or self-employed;

(vi) has non-zero monthly income in current year; and

(vii) is a household head in the city.[6]

Since we have to predict each migrant's potential earnings in each city, we need to run a separate earnings regression for each single city. For sample size reasons, we drop all cities with fewer than 30 rural-urban migrants. Raising this cutoff point higher will allow us to estimate the income equation more precisely for the smallest cities in the sample. However, it also means that the sample size will be smaller for the city-level regressions in the second stage. We decide to use 30 migrants as the cutoff point because we find that, with this sample size, we can still estimate the income equation with a reasonable precision. In our sensitivity analysis, we will check whether this arbitrary cutoff significantly affects our main results.

In addition, we have to drop seven cities for which the instrumental variable is missing. With all these restrictions on the data sample, we have a total of 95 cities that will be used for our baseline regressions. Among the cities that were screened out of the sample, most are so small that they have relatively few rural-urban migrants. Whereas we are dropping 70 percent of the prefecture-level cities (222 out of 317) in the survey data, we have only excluded 9.97 percent of the rural-urban migrants (2,690 out of 26,986) from our analysis.

Some descriptive statistics are shown in Table 1. For comparison purposes, we have also included descriptive statistics of local urban workers in these cities. Migrants tend to be younger; the average migrant is 32.6 years old, compared to the 40.3 years of age for the average urban worker. Perhaps because they are younger, a larger share of rural-urban migrants are unmarried (23.3 vs. 10.0 percent). The average migrant is less educated, with 9.1 years of schooling compared to the average local resident's 12.4 years of schooling.

---

[5]As far as we know, all academic researchers who have access to this data only have one fifth of the sample.

[6]Following common practice in the literature, we focus on household heads in our empirical analysis, assuming that they are the decision makers. A young migrant might live with his or her parents back in the home village and thus is not considered a household head in the village. Here we consider such a migrant a household head if he or she lives alone in the city.

Table 1: Descriptive statistics for migrant and urban household heads

Variables	Rural-Urban Migrants		Urban Workers	
	Mean	Std. Dev.	Mean	Std. Dev.
Age	32.63	8.404	40.25	9.124
Age < 30	0.438	0.496	0.165	0.371
Female	0.210	0.407	0.219	0.414
Unmarried	0.233	0.423	0.100	0.300
Years of schooling	9.072	2.416	12.446	3.012
Education levels				
Elementary school or below	0.189	0.392	0.036	0.187
Middle school	0.593	0.491	0.245	0.430
High school or above	0.218	0.413	0.719	0.450
Self-employed	0.254	0.435	0.089	0.284
Monthly income (yuan)	1,129.8	785.3	1,678.3	1,517.4
No. of observations	24,296		62,223	

Statistics in this table are based on the sample of migrant and urban household heads between 20 and 60 years old. Observations in 95 cities with at least 30 migrants are included in this calculation.

Migrants are much more likely to be self-employed and have a much lower monthly income than urban workers. Only a little over 20 percent of migrants or urban workers are women. This is because our analysis focuses on household heads only and there are fewer female-headed households.

For rural-urban migrants, we also examine where they come from and where they currently reside, which is shown in Table 2. A few facts are worth noting. First, the South and the East are the two leading destination regions. The Pearl River Delta area is in the South; the Yangtze River Delta area is in the East. These two areas are the major manufacturing hubs in China, where the labor-intensive industries rely heavily on migrant workers. Second, the Central region, although it has supplied far more migrants than any other region, absorbs only a small number of migrants. In fact, it is the smallest destination region, even slightly behind the economically backward Northwest region. Third, short-distance migration is more common than long-distance migration. For most destination regions, the majority of the migrants come from within the region.[7] Indeed, the East is the only region where the largest share of migrants is not from within the region (but from the Central region).

In addition, we also use several ancillary data sources to construct some other variables for this study. The first one is migration distance. From the population survey, we know the home and destination prefectures of each migrant. We use the latitude-longitude coordinates of each prefecture to calculate the "great-circle distance" (on the surface of the Earth) between the home and destination prefectures .[8]

---

[7]Zhang and Zhao (2013) show that rural-urban migrants in China prefer to stay close to their home villages. They attempt to measure the amount of income these migrants are willing to give up in order to stay closer to home.

[8]We calculate this distance using the Haversine formula (Sinnott, 1984). Let $(lat_1, long_1)$ and $(lat_2, long_2)$

Table 2: Migration flows within and across regions

Origin regions	Destination regions							Row total
	North	Northeast	East	Central	South	Northwest	Southwest	
North	1,576	61	185	4	98	16	11	1,951
Northeast	288	790	38	1	61	7	6	1,191
East	230	32	2,662	19	262	30	50	3,285
Central	847	64	3,612	566	3,419	121	59	8,688
South	23	2	50	5	3,773	0	5	3,858
Northwest	98	32	100	4	148	397	21	800
Southwest	279	13	1,473	34	1,792	67	865	4,523
Column total	3,341	994	8,120	633	9,553	638	1,017	24,296

Statistics in this table are based on the sample of migrant and urban household heads between 20 and 60 years old. Observations in 95 cities with at least 30 migrant household heads are included in this calculation. The number in each cell is the total number of migrants who moved from the origin (row) region to the destination (column) region. Following cultural geographers, we divide China into seven regions as follows: North (Beijing, Tianjin, Hebei, Shandong, Shanxi); Northeast (Liaoning, Jilin, Heilongjiang, Neimenggu); East (Shanghai, Jiangsu, Zhejiang, Fujian); Central (Henan, Anhui, Jiangxi, Hubei, Hunan); South (Guangdong, Guangxi, Hainan); Northwest (Shaanxi, Gansu, Ningxia, Xinjiang); Southwest (Sichuan, Chongqing, Guizhou, Yunnan, Qinghai, Xizang).

For city-level regressions, we collect information on city characteristics in 2005 from the *Urban Statistical Yearbook of China*. There is one city-amenity variable, average January temperature, which we think is important but is not available from the yearbook. We hand-collect this data from the online China Meteorological Data Sharing Service System.[9] To measure city population size, we use the data from the One-Percent Population Survey to calculate the total number of residents living in each city, counting both the regular residents with a local *hukou* and the rural-urban migrants in the city. To instrument for city population size, we use a long lag of this variable, which is from the 1953 census, the

be the latitude-longitude coordinates of two locations, then the shortest distance between them over the Earth's surface, d, is given by:

$$
\begin{aligned}
\Delta lat &= lat_2 - lat_1 \\
\Delta long &= long_2 - long_1 \\
a &= \left[\sin\left(\frac{\Delta lat}{2}\right)\right]^2 + \cos(lat_1) \cdot \cos(lat_2) \cdot \left[\sin\left(\frac{\Delta long}{2}\right)\right]^2 \\
c &= 2 \cdot \text{atan2}\left(\sqrt{a}, \sqrt{1-a}\right) \\
d &= r \cdot c
\end{aligned}
$$

where $r = 6{,}371\,km$ is the mean value of the Earth's radius. Note that angles need to be in radians in the calculation.

[9]The website, http://cdc.cma.gov.cn (accessed February 22, 2012), is maintained by the National Meteorological Information Center at the China Meteorological Administration. They collected data from 134 meteorological stations throughout China and calculated the 1971-2000 average monthly temperature at each station. For each city in our sample, we use the average temperature from the nearest meteorological station.

first national census in modern China.[10]

5 Estimation

We present estimation results in this section.

5.1 Potential earnings

Our first task is to predict $\ln \hat{I}_{ij}$, each migrant i's potential earnings in each city j. A naive method would be to run a city-specific OLS regression of income on individual characteristics for every city and then predict each migrant's income in each city using the estimated income equation for that city. Indeed, Timmins (2007) used this method. However, such simple OLS regressions are likely to produce biased estimates because of sorting across cities. For example, some migrants choose to move to Shanghai perhaps because they are ambitious and have high hopes for the future. Such unobserved characteristics may be correlated with observed migrant characteristics such as education. If we ignore this self-selection problem, the education coefficient in the income equation for Shanghai will be biased, and thus we cannot accurately predict potential earnings in Shanghai for those migrants who are not currently working in the city. To correct for this kind of selection biases, we follow a semi-parametric approach, a method developed by Dahl (2002) and used by Bayer et al. (2009) to predict earnings for internal migrants in the U.S.[11]

To demonstrate Dahl's method, consider the following empirical model

$$\ln I_{ij} = Z_i \gamma_j + \mu_{ij}, \tag{11}$$

where $\ln I_{ij}$ is log income for individual i in city j; Z_i a vector of individual characteristics; and μ_{ij} the error term. Further assume that $\ln I_{ij}$ is observed if and only if individual i chooses city j among a total of J alternatives, which happens when a latent variable (e.g., utility) is maximized in j.

Dahl (2002) shows that one can obtain a consistent estimate of γ_j by the regression

$$\ln I_{ij} = Z_i \gamma_j + \psi\left(P_{i1}, ..., P_{iJ}\right) + e_{ij},$$

where P_{ij} is the probability of i choosing j and $\psi\left(\cdot\right)$ an unknown function that gives the

---

[10]Some cities in our sample were not prefecture-level cities in 1953; at that time they were the major towns in their rural counties. The 1953 population size for these towns are not available, for which we use the 1953 county population instead.

[11]When estimating income for migrants, researchers have long recognized the self-selection problem. See, for example, Nakosteen and Zimmer (1980), Robinson and Tomes (1982), and Falaris (1987). Falaris actually considers self-selection in a multiple choice migration model, a situation similar to ours. He uses an estimator proposed by Lee (1983). We decide to use the more recent semi-parametric approach developed by Dahl (2002) primarily because Monte Carlo simulations suggest that Dahl's method is preferred to Lee's (Bourguignon et al., 2007).

conditional mean of the error term in equation (11), $E(\mu_{ik}|\cdot)$. Dahl (2002) introduces a "single-index sufficiency assumption" which assumes that the probability of the first-best choice is the only information needed for estimating the conditional mean. This dramatically reduces the dimension of the correction function ψ and the above estimation equation becomes

$$\ln I_{ij} = Z_i \gamma_j + \widetilde{\psi}(P_{ij}) + e_{ij}.$$

Since i has indeed chosen city j, Dahl (2002) proposes to estimate P_{ij} nonparametrically based on actual migration flows. The unknown function $\tilde{\psi}$ can be approximated by linear expansions.

Following this approach, for each destination city j, we use the information about the migrants who currently reside in this city to estimate an equation for log income. The key to implementing Dahl's method is to nonparametrically estimate the probability of each individual migrating to her city. We first divide all the individuals into different "cells" based on home region, education level, and age. Following cultural geographers, we divide China into seven different regions: North, Northeast, East, Central, South, Northwest, and Southwest. Within each of the seven home regions, individuals are divided into a "high-education" group (with more than 9 years of schooling) and a "low-education" group (with no more than 9 years of schooling). They are then further categorized into a "young" group (age ≤ 30) and an "old" group (age > 30). Thus we have classified all the migrants into 28 different cells.[12] For each individual i in city j, we find the cell to which she belongs. The estimated probability of i choosing j, $\hat{P}_{ij}$, is simply calculated as the fraction of all the individuals in that cell who migrated to city j.

For each city j, we regress log income on a vector of individual characteristics and a second degree polynomial of $\hat{P}_{ij}$:

$$log\_income = a + b_1 * age + b_2 * age\_squared + b_3 * gender + b_4 * schooling$$
$$+ c_1 * \hat{P}_{ij} + c_2 * \hat{P}_{ij}^2 + e_{ij}.$$

We then use this estimated equation to predict $\ln \hat{I}_{mj}$ for every migrant m in our sample. Two notes are in order regarding this procedure. First, we used the information on age, schooling, and home region to predict migration probability $\hat{P}_{ij}$. Since both age and schooling are also included in the income equation here, identification requires that home region should be excluded from the income equation. That is, we are assuming here that once individual characteristics are controlled, a migrant's birth place does not help predict earnings at any migration destination. Second, we add $\hat{P}_{ij}$ and its square term to the regression only for consistently estimating $b_1 - b_4$. We do not need them when predicting income.

---

[12]There is a tradeoff between having more cells and the precision of estimated migration probability. Because each individual can choose among more than 100 different destination cities, we need a reasonably large number of individuals in each cell in order to have a good estimate of the probability. For this reason, we cannot divide our sample into too many cells.

Table 3: Regression results from the conditional logit model

Variable | coefficient name	Coefficient	Standard Error	z-statistic
Utility from income			
Log income | α	0.539	0.081	6.695
Migration cost			
Log migration distance | π_D	-0.964	0.013	-71.73
Adjacent province | π_1	-2.485	0.032	-78.57
Non-adjacent province | π_2	-3.503	0.043	-82.15
City fixed effects		Included	
Wald chi2(98) p-value		0.0000	
Number of cities		95	
Number of observations		2,308,120	

Only cities with at least 30 migrant household heads are included in this regression. The number of observations equals the number of migrants (24,296) multiplied by the number of destination cities (95).

5.2 City fixed effects

With the predicted income for every migrant in every destination city, we now estimate the conditional logit model by maximizing the likelihood function given by equation (9). Note that only log income, migration distance variables, and city fixed effects are included in this regression. The results are in Table 3. As expected, the utility from income is positive, and it is very precisely estimated. Also consistent with our expectation, migration distance causes disutility. In addition, moving to an adjacent province, compared to staying within the home province, is associated with a decline in utility. Moving further away incurs an even larger loss in utility.

Setting the city fixed effect for Beijing to be zero, we have estimated a θ_j for each city j. It represents the average migrant's willingness to pay for living in each city, controlling for potential earnings and migration costs. Another way to interpret this city fixed effect is to view it as a "quality of life" measure, with a higher θ_j representing a better quality of life. In Table 4, we list the top 20 cities with best quality of life. On top of the list are Shanghai, Shenzhen, Beijing, and Guangzhou. These are the cities of both political and economic importance. They are usually considered the face of modern China. They are also the cities in which typical Chinese aspire to live. Except Beijing, Tianjin, Shenyang, and Dalian, all other top cities are in the East or the South, the two regions with the most prosperous regional economies. Overall, the list does seem to be consistent with our prior knowledge of cities with high qualities of life in China.

5.3 City size and urban amenities

Next we present results from our second-step regression. Here we regress the city fixed effects on a set of observed city characteristics, focusing on city population size as the key explanatory variable. We are essentially estimating equation (10) except that we add a

Table 4: Top twenty cities ranked by rural-urban migrants' willingness to pay

Rank	City	Value of $\hat{\theta}_j$
1	Shanghai	0.7551
2	Shenzhen	0.2851
3	Beijing	0.0000
4	Guangzhou	-0.1157
5	Foshan	-0.2891
6	Ningbo	-0.3734
7	Wenzhou	-0.6746
8	Tianjin	-0.7104
9	Quanzhou	-0.9479
10	Shenyang	-1.0716
11	Hangzhou	-1.1477
12	Jinhua	-1.1529
13	Xiamen	-1.1701
14	Suzhou	-1.3940
15	Jiaxing	-1.4136
16	Fuzhou	-1.5250
17	Dalian	-1.6092
18	Putian	-1.6180
19	Nanjing	-1.6494
20	Dongguan	-1.7896

Notes: The value of $\hat{\theta}_j$ is estimated in the regression presented in Table 3.

constant term, $\overline{\theta}$, to account for the utility derived from the average city in the sample:

$$\theta_j = \overline{\theta} + \beta_S^* \ln S_j + \sum_{k=1}^{K} \beta_k^* \ln X_{jk} + \xi_j^*. \tag{12}$$

As argued above, the main reason we take the two-step approach to estimating our model is the concern of a potential omitted-variables problem. That is, unobserved city characteristics (e.g., migrant-friendliness and pro-growth local economic policies) may affect both the typical migrant's utility and city population, which would bias the estimation of β_S^*, our key parameter of interest. Using the two-step method, we will have a simple linear regression at the second step, which allows us to adopt two standard strategies to deal with the omitted-variables problem. First, we add region dummies in our regression, attempting to identify β_S^* using only within-region variations. Since we divide China into seven regions, omitted-variable concerns mostly arise from cross-region differences. Controlling for region fixed effects should help mitigate the potential omitted-variable bias.

Second, and more importantly, we take the instrumental variables approach, which with valid instruments can deal with not only omitted-variables but also simultaneity and measurement-errors problems. Our analysis at the second step focuses primarily on the coefficient of city population. Following the tradition in the literature, we use long lags of

city population as the instrument. More specifically, we use lagged values of city population from the 1953 census, the first national census in modern China. Using the lagged variable as an instrument is based on two beliefs. First, there is some persistence in city population, so that the lagged variable is correlated with its current value and thus satisfies the relevance condition for an instrument. This condition is, of course, verifiable with data. Second, historical conditions are dramatically different from today and therefore not directly responsible for today's outcome, which is the exogeneity requirement for a valid instrument. As with any instrumental variable, this second condition is an assumption that cannot be directly tested. We believe that this exogeneity condition is likely to hold in our case. In 1953, China was a backward economy that had a small urban sector. From 1953 to 2005, the country's population almost tripled; its urban population share increased from 13 to 43 percent; a planned economy was established in the first half of this period and was gradually replaced by a market-oriented system in the second half. A series of radical reforms, both political and economic, were implemented over this period of time, which dramatically redefined the landscape of urban China. Thus it seems safe to say that if some unobserved factors or events had a major effect on both population growth in Chinese cities and the utility levels of urban residents in 2005, they must have occurred after 1953. Therefore, it is reasonable to treat the 1953 city population as being exogenous and excludable from our city-level regression.

In Table 5, we present descriptive statistics of variables used in our second-step regression. The city characteristics included in the regression are: population, population density, per capita GDP, unemployment rate, number of large industrial enterprises, share of domestic firms among large industrial enterprises, per capita elementary schools, per capita paved road area, industrial particulate emission, and average January temperature. All city characteristics are measured in log terms, except the average January temperature that has negative values. We will include the square term of the temperature variable in our regressions to allow for a possible nonlinear relationship.

Table 6 presents the correlation matrix for all the variables used for city-level regressions. The first two columns are particularly informative. In the first column, we see that as expected the estimated city fixed effects are indeed positively correlated with log population. That is, migrants are willing to give up some income in order to live in larger cities. Other correlation coefficients in the first column suggest that migrants prefer higher per capita GDP, more large industrial enterprises, and lower share of domestic firms (i.e., higher share of foreign owned firms); they also prefer more paved roads and lower industrial emission of air pollutants. All these make sense, suggesting that the estimated city fixed effects variable is indeed a good measure of the value of urban amenities. Migrants also appear to prefer high-density cities, perhaps because high-density cities tend to have more urban amenities and better public facilities (which can be supplied at lower average costs in high-density areas). There is only one significant coefficient in column 1 that does not seem immediately

Table 5: Descriptive statistics of city characteristics

| Variables	Description	Mean	Std. Dev.	Min.	Max.
Theta	City fixed effects θ_j	-2.942	1.500	-6.415	0.755
Pop	Log(population, 10,000 persons)	5.077	0.866	3.358	7.247
Pop1953	Log(1953 population, 10,000 persons)	3.362	1.113	1.137	6.430
PopDen	Log(population density, persons/square km)	6.878	1.026	3.523	8.842
GDP	Log(per capita GDP)	10.161	0.635	8.471	12.17
UnempRat	Log(unemployment rate)	-3.597	0.582	-5.109	-2.510
LargeFirm	Log(no. of large industrial enterprises per 10,000 persons)	1.125	0.878	-0.512	4.163
DomShare	Log(share of domestic firms in large industrial enterprises)	-0.296	0.286	-1.516	0.000
Schools	Log(no. of elementary schools per 10,000 persons)	0.639	0.543	-0.754	2.057
PavedRd	Log(per capita paved road area, square meter)	2.159	0.545	0.713	3.689
Emission	Log(industrial particulate emission, 1,000kg/10,000 persons)	4.716	1.407	-0.240	7.565
JanTemp	Average January temperature, 1971-2000	3.260	9.417	-21.70	19.00

Statistics are calculated using variables for 95 cities, which are included in the baseline regressions.

Table 6: Correlation matrix of city characteristics

	Theta	Pop	Pop1953	PopDen	GDP	UnempRat	LargeFirm	DomShare	Schools	PavedRd	EmissionJanTemp
Theta	1.000										
Pop	.596*	1.000									
Pop1953	.266*	.563*	1.000								
PopDen	.283*	.519*	.307*	1.000							
GDP	.201*	.205*	.325*	-.176	1.000						
UnempRat	-.118	.124	.069	.159	-.101	1.000					
LargeFirm	.488*	.251*	.203*	.096	.632*	-.190	1.000				
DomShare	-.383*	-.102	-.145	-.222*	-.142	.351*	-.442*	1.000			
Schools	-.397*	-.448*	-.136	-.182	-.360*	-.067	-.479*	.139	1.000		
PavedRd	.300*	.298*	-.072	.134	.315*	-.027	.345*	-.256*	-.567*	1.000	
Emission	-.346*	-.155	.185	-.304*	.082	.168	-.134	.395*	.031	-.008	1.000
JanTemp	.040	-.114	.037	.214*	-.197	-.194	.251*	-.409*	.098	-.207*	-.502* 1.000

Correlations are calculated using variables for 95 cities, which are included in the baseline regressions. See variable descriptions in Table 5. * denotes p < 0.05.

Figure 1: The relationship between city fixed effects ($\hat{\theta}_j$) and log population
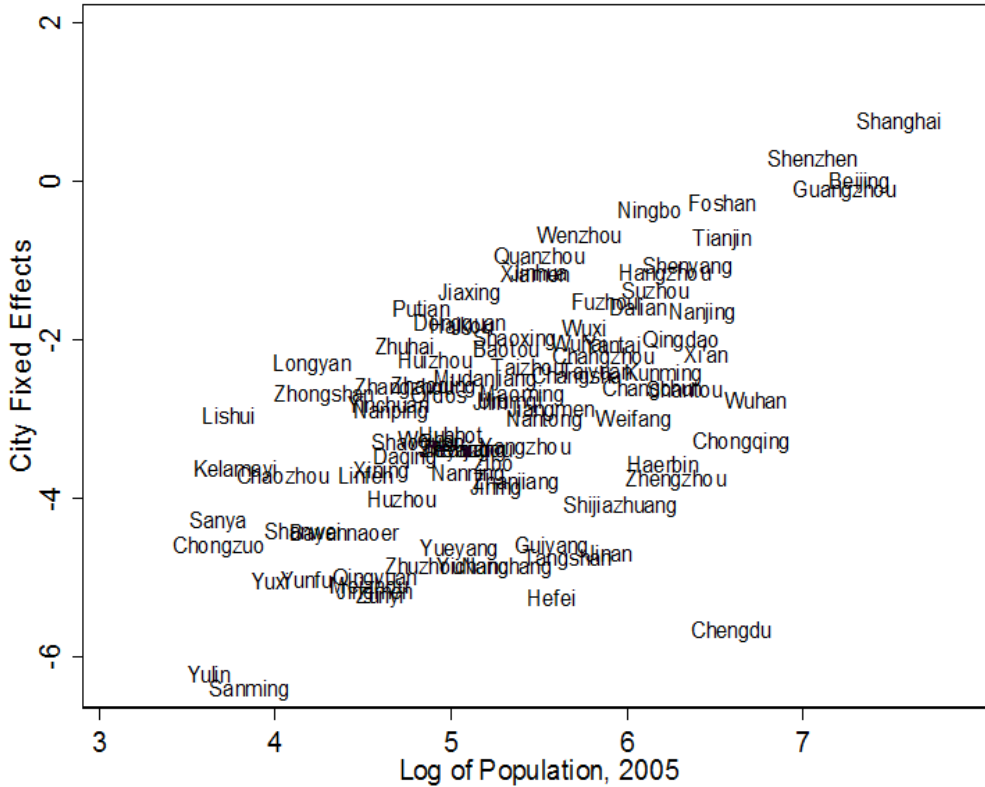
obvious: migrants prefer fewer elementary schools (per 10,000 residents). One possible reason is that cities with fewer elementary schools tend to have larger schools, which are generally of higher quality.

It is important to note that population size is correlated with many observed city characteristics, as shown in column 2 of Table 6. Larger cities have higher population density, higher per capita GDP, and more large industrial enterprises. Larger cities also have fewer (and thus larger) elementary schools, more paved roads, and lower industrial emissions. This correlation between population size and other city characteristics is exactly the reason why we should examine whether migrants prefer larger cities per se or they prefer larger cities for their observed characteristics. And finally, current population size is highly correlated with its lagged value in 1953, a necessary condition for the latter to be a valid instrumental variable.

5.3.1 City size

We start by examining migrants' utility derived from a single city characteristic, population size. Figure 1 plots city fixed effects ($\hat{\theta}_j$) versus log population size in 2005. We see a clear positive relationship: migrants derive higher utilities from larger cities.

Table 7: Effects of city size on utility from single-variable regressions

Dependent Variable = city fixed effect ($\hat{\theta}_j$)

Variables	(1) OLS	(2) OLS	(3) IV	(4) IV
Log (population, 10,000 persons)	1.268*** (0.123)	1.192*** (0.114)	0.895*** (0.209)	1.056*** (0.133)
Constant	Yes	Yes	Yes	Yes
Region dummies	No	Yes	No	Yes
Adjusted R^2	0.536	0.878	0.489	0.774
No. of observations	95	95	95	95

In IV specifications, log city population is instrumented using its lagged In our n 1953. Standard errors are in parentheses. *** statistically significant at the 1% level. Regressions are weighted by the number of rural-urban migrants in each city.

Table 7 presents results from regressing city fixed effects on the single right-hand side variable, log city population. Columns (1)-(2) are OLS results. The only difference between the two specifications is that column (1) does not control for region dummies but column (2) does. Columns (3)-(4) are parallel to (1)-(2), except that we instrument for log city population using its lagged values in 1953.

Estimated coefficients in Table 7 indicate that rural-urban migrants indeed derive more utilities from larger cities. The preferred estimate, from column (4) that uses the IV specification and controls for region fixed effects, is 1.056. From Table 3, we know that the coefficient of log income in the indirect utility function is 0.539. Their ratio, 1.056/0.539 = 1.96, implies that a typical rural-urban migrant is willing to give up 1.96 percent of her income in order to live in a city with a one-percent higher population. This is clearly a very high elasticity.

Comparing columns (1)-(2) to their corresponding IV specifications in columns (3)-(4), we find that the IV coefficients are lower than the OLS coefficients, especially when we do not control for region fixed effects. If unobserved city characteristics or policies attract migrants to a city, they are likely to increase individual migrant's utilities. Thus omitted variables are indeed expected to cause upward biases in OLS coefficients. The results in Table 3 suggest that such biases do exist and seem to be more serious when region fixed effects are not controlled.

5.3.2 Controlling for city characteristics

We next experiment with different specifications by adding city characteristics to the regression. We have again tried both OLS and IV regressions. In all specifications, OLS and IV coefficients are similar. Therefore, to conserve space, our presentation here focuses on the results from IV regressions only. Since there are many city characteristics, we divide

21

them into three groups. The first group concerns the city's economic conditions, including per capita GDP, unemployment rate, number of large industrial enterprises per 10,000 residents, share of domestic firms in large industrial enterprises, all in logs. The second group is about educational facilities and infrastructure, including number of elementary schools per 10,000 residents and per capita paved road area, also in logs. The third group characterizes the city's living environment, including log population density, log industrial particulate emission, average January temperature and its square term.

Our goal here is to examine whether individuals are willing to pay for large cities only because such cities have desirable characteristics. That is, if adding some controls substantially reduces the magnitude of the population size coefficient, then we take it as evidence that people prefer those city characteristics that come with a large population, rather than the large population itself. We add each group of control variables to the regression separately, and finally include all of them in a single regression. Table 8 presents regression results without controlling for region fixed effects; Table 9 presents results from parallel regressions controlling for region fixed effects.

The first thing to notice is that in all of these regressions, log city population still has positive coefficients and they are all statistically significant. That is, even if we control for so many city characteristics, large cities still appear to be attractive to rural-urban migrants.

Results in Tables 8 and 9 tell a similar story: The magnitude of the coefficients obtained from multivariate regressions are comparable to those from the single-variable regressions (reproduced as column (1) in both tables), and some are even larger. In other words, people strongly prefer larger cities, and this result cannot be explained away by the city characteristics included in these regressions. The coefficient of population size is always larger when region fixed effects are controlled, suggesting a negative correlation between unobserved regional differences and city sizes. Our preferred estimate of β_S^*, from the last specification in Table 9 that controls for all city characteristics as well as the region dummies, is 0.929. Together with the estimated α in Table 3 (0.539), it implies that the typical migrant is willing to give up 1.72 percent of monthly earnings for a one percent increase in city population. This is still a very high elasticity.

The results on some of the control variables are also of interest. Coefficients in both Tables 8 and 9 indicate that rural-urban migrants prefer cities with more large industrial enterprises, perhaps because such cities offer more job opportunities. Another consistently significant result in both tables is that migrants prefer cities with less air pollution. These regressions also suggest that migrants like cities with lower unemployment rates and more foreign-owned large industrial enterprises, which makes sense. One result appears to be counter-intuitive at first sight: per capita GDP has a negative coefficient (although not always statistically significant). This is not puzzling once we realize that the coefficient estimated here captures the overall effects of GDP on utility. Whereas higher GDP itself may be a good thing, it may well become unattractive when migrants factor in the higher

Table 8: Effects of city size on utility, controlling for city characteristics

Dependent Variable = estimated city fixed effect ($\hat{\theta}_j$)

Variables	(1) IV	(2) IV	(3) IV	(4) IV	(5) IV
Log(population, 10,000 persons)	0.895***	1.047***	0.696***	0.728***	0.678***
	(0.209)	(0.124)	(0.250)	(0.265)	(0.260)
Log(per capita GDP)		-0.529***			-0.577***
		(0.175)			(0.207)
Log(unemployment rate)		-0.141			-0.107
		(0.150)			(0.170)
Log(no. of large industrial enterprises per 10,000 persons)		0.900***			0.876***
		(0.144)			(0.173)
Log(share of domestic firms in large industrial enterprises)		-0.446			-0.453
		(0.309)			(0.372)
Log(no. of elementary schools per 10,000 persons)			-0.376		-0.550**
			(0.235)		(0.271)
Log(per capita paved road area, square meter)			0.618***		-0.301
			(0.218)		(0.184)
Log(population density, persons/square km)				0.115	0.104
				(0.184)	(0.174)
Log(industrial particulate emission, 1,000kg/10,000 persons)				-0.416***	-0.363***
				(0.166)	(0.131)
Average January temperature, 1971-2000				0.016	-0.035**
				(0.018)	(0.014)
Average January temperature squared				-0.000	0.002
				(0.001)	(0.001)
Constant	Yes	Yes	Yes	Yes	Yes
Region dummies	No	No	No	No	No
Adjusted R^2	0.489	0.734	0.514	0.594	0.749
Number of observations	95	95	95	95	95

In all specifications, log city population is instrumented using its lagged value in 1953. Standard errors are in parentheses. * statistically significant at the 10% level; ** statistically significant at the 5% level; *** statistically significant at the 1% level. Regressions are weighted by the number of rural-urban migrants in each city.

Table 9: Effects of city size on utility, controlling for city characteristics and region dummies

Dependent Variable = estimated city fixed effect ($\hat{\theta}_j$)

Variables	(1) IV	(2) IV	(3) IV	(4) IV	(5) IV
Log(population, 10,000 persons)	1.056***	1.166***	1.087***	0.874***	0.929***
	(0.133)	(0.100)	(0.115)	(0.188)	(0.228)
Log(per capita GDP)		-0.232			-0.196
		(0.154)			(0.171)
Log(unemployment rate)		-0.397***			-0.375***
		(0.144)			(0.142)
Log(no. of large industrial enterprises per 10,000 persons)		0.410***			0.422***
		(0.156)			(0.151)
Log(share of domestic firms in large industrial enterprises)		-0.232			-0.310
		(0.261)			(0.316)
Log(no. of elementary schools per 10,000 persons)			0.094		-0.141
			(0.200)		(0.265)
Log(per capita paved road area, square meter)			0.093		-0.206
			(0.154)		(.149)
Log(population density, persons/square km)				-0.058	-0.067
				(0.136)	(0.149)
Log(industrial particulate emission, 1,000kg/10,000 persons)				-0.358***	-0.318***
				(0.110)	(0.107)
Average January temperature, 1971-2000				0.034	0.017
				(0.030)	(0.029)
Average January temperature squared				-0.001	-0.000
				(0.002)	(0.002)
Constant	Yes	Yes	Yes	Yes	Yes
Region dummies	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.774	0.824	0.772	0.794	0.836
Number of observations	95	95	95	95	95

In all specifications, log city population is instrumented using its lagged value in 1953. Standard errors are in parentheses. * statistically significant at the 10% level; ** statistically significant at the 5% level; *** statistically significant at the 1% level. Regressions are weighted by the number of rural-urban migrants in each city.

Table 10: Results from sensitivity analysis

Alternative samples or specifications	α	β_S^*(or β_S)	$\frac{\beta_S^*}{\alpha}$ (or $\frac{\beta_S}{\alpha}$)
A. Baseline results; use the 30-migrant cutoff	0.539*** (0.076)	0.929*** (0.228)	1.72
B. Use the 40-migrant cutoff	0.545*** (0.091)	0.869*** (0.262)	1.59
C. Use the 20-migrant cutoff	0.319*** (0.067)	0.843*** (0.310)	2.64
D. D.V. = city fixed effect + 0.28* Log(quality-adjusted housing price)	0.539*** (0.076)	0.946*** (0.231)	1.755
E. D.V. = city fixed effect + 0.22* Log(quality-adjusted housing price)	0.539*** (0.076)	0.943*** (0.230)	1.750
F. Allow for heterogeneous preferences	0.346*** (0.083)	0.616*** (0.214)	1.78

Baseline results are from Table 3 and column (5) of Table 9. In all specifications, log city population is instrumented using its lagged value in 1953. Standard errors are in parentheses. * statistically significant at the 10% level; ** statistically significant at the 5% level; *** statistically significant at the 1% level. Regressions are weighted by the number of rural-urban migrants in each city.

housing prices they have to pay in high-GDP cities.

5.4 Sensitivity analysis

We next explore the sensitivity of our baseline results to alternative specifications.

Estimate income using alternative samples

A key step in our empirical estimation is to predict income for each migrant in each potential destination city. This is very demanding of data because it requires a reasonably large migrant sample in each city included in our city-level regression. We are facing the following tradeoff: On the one hand, we would like to have a large sample for the city level regression, and thus prefer to include many cities even if some of them are relatively small. On the other hand, when we include more small cities, the number of observed migrants in each of the smaller cities will be quite low; and thus expected earnings in these cities will have to be estimated based on very limited information. In our baseline regressions, we decide to use a 30-migrant cutoff: A city is included in our analysis sample only if at least 30 migrant household heads are observed in the city. Here we explore whether our main results are driven by this arbitrary cutoff.

We try two alternative criteria for selecting our estimation sample. First, we increase the cutoff to 40 migrants, which reduces the sample size from 95 to 77 for the city-level regression. Second, we lower the cutoff to 20 migrants, which increases the sample size from 95 to 126 for the city-level regression.

In Table 10, we first reproduce the baseline estimates for α and β_S^* in row A. The estimate for α is from Table 3 and β_S^* is our preferred estimate from the specification that controls for all city characteristics as well as region fixed effects (Table 9, column (5)). In rows B-C, we present estimates from the same specification except that the expected migrant income is estimated in alternative samples as just described. In each case, α and β_S^* are still positive and statistically significant. Since the coefficients are rescaled in each estimation with a different sample, we cannot directly compare the estimates of α or β_S^* across different rows. However, different estimates of the elasticity $\frac{\beta_S^*}{\alpha}$ can be compared, so we show this ratio in the last column. Using the 40-migrant cutoff gives an estimate of $\frac{\beta_S^*}{\alpha}$ somewhat smaller than the baseline result, 1.59 vs. 1.72. The 20-migrant cutoff leads to an estimate of 2.64, considerably larger than the baseline estimate. Results in rows A-C seem to suggest that using a smaller sample of cities (i.e., excluding many small cities) tends to give a smaller estimate of the elasticity. Overall, the qualitative results remain the same regardless of which sample of cities is used.

Deal with housing price explicitly

Recall that $\beta_S^* \equiv \beta_S - \alpha_H \lambda_s$. While migrants derive utility directly from larger population sizes (β_S), they also dislike higher price levels that go with certain population sizes ($-\alpha_H \lambda_s$). That is, our estimate of β_S^* captures the net effect of population size on utility. Here we go one step further and try to identify β_S, the gross preference for larger population size. From the model, a positive β_S would imply that people prefer larger cities. However, so far we have only shown a positive β_S^*. Notice that $\beta_S > 0$ and $\beta_S^* > 0$ are not equivalent; we could have $\beta_S \leq 0$ and $\beta_S^* > 0$ if price level is lower in larger cities (i.e., $\lambda_s < 0$). Our exercise here is to check whether $\beta_S > 0$.

To proceed, we assume that the price of the nontradable good in our model, p_j, is entirely driven by the price of housing services in city j. Instead of assuming that $\ln p_j$ is a function of city characteristics (as in equation (3)), we now simply treat it as housing price and deal with it explicitly. We go back to equation (2) and redefine the average utility derived from city j as

$$\theta_j = -\alpha_H \ln p_j + \beta_S \ln S_j + \sum_{k=1}^{K} \beta_k \ln X_{jk} + \xi_j.$$

Notice that housing price $\ln p_j$ may be correlated with both observed and unobserved city characteristics ($\ln X_{jk}$ and ξ_j), which could potentially bias our estimate of β_S. To address this problem, we follow the strategy of Bayer et al. (2009) by moving $-\alpha_H \ln p_j$ to the left hand side and estimating β_S from the following equation:

$$\theta_j + \alpha_H \ln p_j = \beta_S \ln S_j + \sum_{k=1}^{K} \beta_k \ln X_{jk} + \xi_j. \tag{13}$$

Here θ_j is still estimated from the first-step conditional logit regression. Parameter α_H is the share of an individual's income spent on housing, which can be calculated from our population survey data.

For a long time in China, individual urban residents could not own housing units. Instead, housing was provided to them by employers at very low costs. Starting in 1998, a reform was implemented to establish a housing market in urban China. However, in the early years, most urban residents were able to buy housing units from their employers at below-market prices, and relatively few families financed the purchase of housing through mortgage loans. For these reasons, information on the payment of home owners in our sample period is hardly useful for estimating housing prices. We therefore focus only on renters to estimate α_H and $\ln p_j$. We estimate α_H using the share of monthly income paid as rent for housing by renters in all cities. One possible approach to estimating $\ln p_j$ is to use an aggregate statistic such as mean or median housing price paid by renters in a city. However, such an aggregate measure is correlated with quality of housing, which is likely correlated with city characteristics, and thus biases the estimates of β_S and β_k. To get around this problem, we estimate $\ln p_j$ from the following housing price regression:

$$\ln R_{ij} = \ln p_j + \Upsilon_i \rho + \nu_{ij}, \tag{14}$$

where $\ln R_{ij}$ is the natural logarithm of individual i's rent payment in city j, $\ln p_j$ a city-specific constant, and Υ_i a vector of i's housing characteristics.[13] That is, $\ln p_j$ is the quality-adjusted housing price in city j.

We try two ways to estimate α_H: one by dividing total rents by total household head's income for all urban households who are renting, which is 0.28; and the other by dividing total rents by total household (instead of just its head's) income, which is 0.22. We then estimate β_S by regressing $\theta_j + 0.28 * \ln p_j$ or $\theta_j + 0.22 * \ln p_j$ on population size and other city characteristics, again instrumenting with lagged population size. The results are in rows D and E of Table 10. Note first that the results are insensitive to how we estimate α_H; the two estimates give almost identical results. Focusing attention on row D, we have a $\beta_S = 0.946$, which is only slightly higher than the baseline estimate $\beta_S^* = 0.929$. This implies that rural-urban migrants indeed prefer larger cities, and this is not because larger cities have lower housing prices. In fact, our results suggest that city size has a positive effect on the quality-adjusted housing price ($\lambda_s = 0.06$).

Allow for heterogeneous preferences for population size

In our baseline estimation, we imposed the assumption that all rural-urban migrants have the same preference for larger cities. That is, β_S^* is constant across individuals. Here

---

[13]In our empirical estimation, we include the following housing characteristics in Υ_i: building age, number of rooms, square meters of living area, story dummies, structure dummies, washroom type dummies, whether is shared with others, whether has a kitchen, whether has a gas connection, whether has a bath, and whether is commercially owned.

we relax this assumption by allowing β_S^* to vary with individual characteristics. More specifically, we denote by Ω_i a vector of four migrant characteristics, including age, gender, years of schooling, and marital status. We interact Ω_i with population size of city j to generate $(\Omega_i \ln S_j)$ and add it to a migrant's indirect utility function (equation 8). The conditional logit estimation in the first step is then performed by maximizing the following likelihood function

$$\widetilde{L} = \prod_i \prod_{j=1}^J \left\{ \frac{\exp\left[\alpha \ln \hat{I}_{ij} + \pi_D \ln D_{ij} + \pi_1 d_{ij}^1 + \pi_2 d_{ij}^2 + (\Omega_i \ln S_j)\delta + \widetilde{\theta}_j\right]}{\sum_{s=1}^J \exp\left[\alpha \ln \hat{I}_{is} + \pi_D \ln D_{is} + \pi_1 d_{is}^1 + \pi_2 d_{is}^2 + (\Omega_i \ln S_s)\delta + \widetilde{\theta}_s\right]} \right\}^{\kappa_{ij}}. \tag{15}$$

Here $\widetilde{\theta}_j$ is still a city fixed effect, but it should be interpreted as the average migrant's willingness to pay for living in city j, assuming that migrants have heterogeneous preferences for city size but homogeneous preferences for other city characteristics.

The estimated coefficients of the interaction terms in this conditional logit regression $(\hat{\delta})$ indicate that better-educated migrants, younger migrants, female migrants, and unmarried migrants have stronger preferences for larger cities. These results are all statistically significant.[14]

We again take the estimated city fixed effects from this conditional logit regression and regress it on city characteristics in the second step. In row F of Table 10, we present the estimates of α and β_S^* from this alternative specification. While both estimates are smaller than those obtained in baseline regressions, the estimated α decreases a little more, leading to a slightly larger income city-size elasticity. Therefore, allowing for heterogeneous preferences for city size does not alter our baseline results.

5.5 Further discussion

One naturally wonders why, if not for observed characteristics, people prefer larger cities. We could come up with several possible explanations.

First, larger cities may provide a better learning environment that allows migrants to accumulate more human capital at a faster rate. There has been evidence that people in cities accumulate human capital and improve their productivity. This productivity gain to a large extent is permanent and portable (Glaeser and Mare, 2001). Recent work has shown that larger cities have a stronger positive effect on residents' productivity (Baum-Snow and Pavan, 2012; De la Roca and Puga, 2012). Therefore, it might be that rural-urban migrants are willing to give up some current income in exchange for higher future income (as they improve their productivity by benefiting from human capital spillovers).

Second, larger cities may offer better life opportunities in the future, either to the

---

[14]We also experimented with adding the interaction term between population size and having children to the regression. In one specification, we added it as an extra interaction term; in the other, we used it to replace the interaction between population size and marital status. In both cases, migrants with children appear to have a weaker preference for larger cities. This is perhaps because currently larger cities tend to make it more difficult for migrant children to enroll in urban public schools.

migrants themselves or to their children. Forward-looking migrants may therefore believe that even if they earn less in a larger city for now, they have a better chance to climb up the social ladder there, or their children will do better in the future. Thus they are willing to give up some earnings today in exchange for a more promising future. In some sense, those migrants living in larger cities are buying a "lottery" that offers a positive although slim chance to win big prizes.[15] The fact that younger and unmarried migrants particularly like larger cities is consistent with our speculation that larger cities provide a better prospect for the migrants, because such people are bound to benefit more from future opportunities.

Third, larger cities may offer a wider variety of consumption goods. In our model, we implicitly assumed that individuals face the same consumption choice set no matter where they live. In reality, this assumption may not hold. In larger cities, there might be a more complete consumption choice set (Glaeser et al., 2001). For example, in a large city like Beijing, one can find all kinds of ethnic food and cultural events; in a smaller city, one may not be able to enjoy such a convenience. With the same amount of income, one could obtain a higher level of utility in a larger city simply because the choice set is larger.

Fourth, larger cities may be more migrant-friendly. Larger cities have always had migrants from different places near and far. Residents in such places tend to have interacted with people of all backgrounds and various origins. They thus tend to be more open-minded and less hostile to migrants. China is a society where personal connections play an important role in everyday life. One would need the support of social and family networks in almost every aspect of life and this is more so in small and close-knit communities. A migrant, as an "outsider," typically has limited support from social and family networks in an urban community. Therefore, they may prefer larger cities where personal connections play a smaller role and so they are less handicapped by the lack of supporting networks.

Whereas it remains unclear what exactly explains the preference for larger cities in China, our findings have immediate policy implications. First, our findings cast doubt on China's long-standing policy of discouraging population growth at large cities. As the Chinese economy continues to grow, an increasingly large rural population will migrate and assimilate into cities. As a former rural resident becomes an urban citizen, the government will presumably take back her land-use rights in her home village and grant her the access to subsidized public facilities and services in a city. Our findings suggest that given the same amount of net subsidy from the government, a rural migrant prefers to live and work in a large city. The current policy in China, which erects barriers around large cities and directs rural migrants to smaller cities, represents a constraint on rural migrants. This kind of policy can only be justified based on externality concerns. However, empirical evidence

---

[15]For these migrants, one particular type of future gain could be a regular urban citizen status. It is widely recognized that the *hukou* system in China is not a sustainable policy and will be reformed sooner or later. It is likely that in the future the government will abolish the *hukou* system and grant rural-urban migrants the access to all subsidized public services in cities. This will be a windfall to the migrants and the windfall is more valuable in larger cities simply because the public services there are better.

does not seem to support the presence of negative externalities of migrants on local urban residents. Au and Henderson (2006a, 2006b) suggest that most Chinese cities are smaller than their efficient sizes. If this is true for large cities, allowing rural migrants to move to undersized large cities will not only benefit these migrants, but also improve the welfare of local residents.[16]

Second, to the extent that urban residents have similar preferences to rural migrants, our findings suggest that the migration restriction in the urban sector also creates inefficiency. Currently, the *hukou* system in China not only prevents rural migrants from freely moving into cities, but also prevents residents with an urban *hukou* from moving across cities. If this cross-city migration restriction is lifted, residents in smaller cities could move to undersized larger cities and greatly improve their welfare, a point that was emphasized by Au and Henderson (2006b).

6   Conclusion

China has a long-standing policy that restricts population growth at large cities but encourages such growth at small and medium-sized cities. At the same time, many other policies favor large cities. We therefore conjecture that quality of life should be higher in larger cities in China. As an implication, we argue that rural-urban migrants should be willing to give up some income in order to live and work in larger cities.

We test this hypothesis using a large-scale population survey conducted in 2005. After controlling for a number of city characteristics, we still find that rural-urban migrants prefer larger cities. Our preferred estimates imply that other things equal, a rural-urban migrant is willing to give up about 1.7 percent of current income in exchange for a one-percent larger city population. We discuss possible factors that may explain this strong preference for large cities. Our findings suggest that China's continuous policy of directing rural migrants away from large cities could cause substantial efficiency losses, especially if many of the large cities are under their efficient sizes, as suggested by earlier research.

References

[1]   Ades, Alberto F. and Edward L. Glaeser (1995). "Trade and Circuses: Explaining Urban Giants," *Quarterly Journal of Economics* 110, 195–227.

---

[16]There has been some research on how rural migrants affect the welfare of regular urban residents in China, focusing mainly on labor market outcomes. Liu and Zhao (2009) find that migrants have negative but small effects on the earnings and employment rates of urban residents. Meng and Zhang (2010) find either small positive or no labor market effects of rural migrants on urban workers. Lu et al. (2012) find that the employment prospects improve as cities become larger, suggesting a positive effect of migrants on urban employment rates. Combes et al. (2013) show that the share of migrants in a city has a large positive effect on the wages of local workers.

[2] Albouy, David (2012). "Are Big Cities Bad Places to Live? Estimating Quality of Life across Metropolitan Areas," unpublished manuscript, Department of Economics, University of Michigan.

[3] Au, Chun-Chung and J. Vernon Henderson (2006a). "Are Chinese Cities Too Small?" *Review of Economic Studies* 73, 549–576.

[4] Au, Chun-Chung and J. Vernon Henderson (2006b). "How Migration Restrictions Limit Agglomeration and Productivity in China," *Journal of Development Economics* 80, 350–388.

[5] Baum-Snow, Nathaniel and Ronni Pavan (2012). "Understanding the City Size Wage Gap," *Review of Economic Studies* 79, 88–127.

[6] Bayer, Patrick, Nathaniel Keohane, and Christopher Timmins (2009). "Migration and Hedonic Valuation: The Case of Air Quality," *Journal of Environmental Economics and Management* 58, 1–14.

[7] Blomquist, Glenn C., Mark C. Berger, and John P. Hoehn (1988). "New Estimates of Quality of Life in Urban Areas," *American Economic Review* 78, 89–107.

[8] Bourguignon, François, Martin Fournier, and Marc Gurgand (2007). "Selection Bias Corrections Based on the Multinomial Logit Model: Monte Carlo Comparisons," *Journal of Economic Surveys* 21, 174–205.

[9] Chan, Kam Wing and Li Zhang (1999). "The Hukou System and Rural-Urban Migration in China: Processes and Changes," *China Quarterly* 160, 818–855.

[10] Combes, Pierre-Philippe, Sylvie Démurger, and Shi Li (2013). "Urbanisation and Migration Externalities in China," CEPR Discussion Paper 9352.

[11] Dahl, Gordon (2002). "Mobility and the Return to Education: Testing a Roy Model with Multiple Markets," *Econometrica* 70, 2367–2420.

[12] De la Roca, Jorge and Diego Puga (2012). "Learning by Working in Big Cities," CEPR discussion paper 9243.

[13] Falaris, Evangelos M. (1987). "A Nested Logit Migration Model with Selectivity," *International Economic Review* 28, 429–443.

[14] Fujita, Masahisa, Tomoya Mori, J. Vernon Henderson, and Yoshitsugu Kanemoto (2004). "Spatial Distribution of Economic Activities in Japan and China," in J.V. Henderson and J.F. Thisse (eds.), *Handbook of Regional and Urban Economics*, Vol. 4, pp. 2911–2977. Elsevier, Amsterdam.

[15] Glaeser, Edward L., Jed Kolko, and Albert Saiz (2001). "Consumer City," *Journal of Economic Geography* 1, 27–50.

[16] Glaeser, Edward L. and David C. Mare (2001). "Cities and Skills," *Journal of Labor Economics* 19, 316–342.

[17] Henderson, J. Vernon (1974). "The Sizes and Types of Cities," *American Economic Review* 64, 640-656.

[18] Henderson, J. Vernon (2005). "Growth of China's Medium-Size Cities," *Brookings-Wharton Papers on Urban Affairs*, 263–295.

[19] Lee, Lung-Fei (1983). "Generalized Econometric Models with Selectivity," *Econometrica* 51, 507–512.

[20] Liu, Xuejun and Yaohui Zhao (2009). "The Impact of Labor Migration on Urban Labor Markets in China," *China Economic Quarterly* 8, 693–710. [in Chinese]

[21] Lu, Jiangyong and Zhigang Tao (2009). "Trends and Determinants of China's Industrial Agglomeration," *Journal of Urban Economics* 65, 167–180.

[22] Lu, Ming, Hong Gao, and Hiroshi Sato (2012). "City Scale and Inclusive Employment," *Social Sciences in China*, Issue 10, 47–66. [in Chinese]

[23] McFadden, Daniel (1974). "Conditional Logit Analysis of Qualitative Choice Behavior," in P. Zarembka, ed., *Frontiers in Econometrics*, Academic Press, New York, 105–142.

[24] McFadden, Daniel (1978). "Modeling the Choice of Residential Location," in A. Karlqvist, L. Lundqvist, F. Snickars, and J. Weibull, eds., *Spatial Interaction Theory and Planning Models*, North-Holland, Amsterdam, 75–96.

[25] Meng, Xin and Dandan Zhang (2010). "Labour Market Impact of Large Scale Internal Migration on Chinese Urban 'Native' Workers," IZA Discussion Paper 5288.

[26] Nakosteen, Robert A. and Michael Zimmer (1980). "Migration and Income: The Question of Self-Selection," *Southern Economic Journal* 46, 840–851.

[27] Roback, Jennifer (1982). "Wages, Rents, and the Quality of Life," *Journal of Political Economy* 90, 1257–1278.

[28] Robinson, Chris and Nigel Tomes (1982). "Self-Selection and Interprovincial Migration in Canada," *Canadian Journal of Economics* 15, 474–502.

[29] Sinnott, Roger W. (1984). "Virtues of the Haversine," *Sky and Telescope* 68, 159.

[30] Timmins, Christopher (2007). "If You Can't Take the Heat, Get Out of the Cerrado... Recovering the Equilibrium Amenity Cost of Non-Marginal Climate Change in Brazil," *Journal of Regional Science* 47, 1–25.

[31] Zhang, Junfu and Zhong Zhao (2013). "Measuring the Income-Distance Tradeoff for Rural-Urban Migrants in China," IZA Discussion Paper 7160.