

Measuring the Income-Distance Tradeoff for Rural-Urban Migrants in China

Junfu Zhang and Zhong Zhao*

May 18, 2011

Abstract

Rural-urban migrants in China appear to prefer nearby destination cities. To gain a better understanding of this phenomenon, we build a simple model in which migrants from rural areas choose among potential destination cities to maximize utility. The distance between a destination city and the individual's home village is explicitly included in the utility function. Using recent survey data, we first estimate an individual's expected income in each potential destination city using a semi-parametric method, controlling for potential self-selection biases. We then estimate the indirect utility function for rural-urban migrants in China based on their migration destination choices. Our findings suggest that to induce an individual to migrate 10 percent further away from home, the wage paid to this migrant has to increase by 15 percent. This elasticity varies very little with distance; it is slightly higher for female than male migrants; it is not affected by the migrant's age, education, or marital status. We explore possible explanations of these results and discuss their policy implications.

Keywords: Income-distance tradeoff, rural-urban migration, *hukou* system, China.

JEL Classification: O15, R12, R23.

[Preliminary draft; comments welcome.]

*Zhang is an assistant professor of economics at Clark University (Tel: 508-793-7247; E-mail: juzhang@clarku.edu). Zhao is a professor of economics at Renmin University of China (Tel: 86-10-8250-2205; E-mail: mr.zhong.zhao@googlemail.com). Both are research fellows at IZA. We thank Wayne Gray and Chih Ming Tan for their thoughtful comments. This paper has also benefited from comments and suggestions by seminar participants at Clark University and the 2nd CIER/IZA Workshop in Bonn, Germany. Xiang Ao, Yue Chen, Kun Guo, Shuyi Lv, and Zehua Sun provided assistance with preparing the migration distance data. Zhong Zhao would like to acknowledge financial support from the Peking University – Lincoln Institute Center for Urban Development.

1 Introduction

China has a residence registration (*hukou*) system, originally designed to control the movement of people within the country. Each family has a registration record, a so-called *hukou*, which specifies the residency status of each individual in the household. It gives a person the right to live and work in a jurisdiction and access local public goods such as public education and health care. Prior to the economic reform, the *hukou* system was strictly enforced. A person with a rural *hukou* could move to a city and work in urban sectors only under very specific situations, which required lengthy and complicated bureaucratic procedures. The quota of such moves was tightly controlled.

Soon after the inception of the economic reform, the rigid *hukou* system was found incompatible with the rapid expansion of the urban economy and the increased demand for cheap labor in urban sectors. Since the mid-1980s, this system has been gradually relaxed and the controls have been weakened (Chan and Zhang, 1999). Most importantly, it has become much easier for a person with a rural *hukou* to obtain a permit to live and work in a city. As a result, China has experienced a massive migration from rural to urban areas in the past three decades. The share of urban population rose from 18 percent in 1978 to 50 percent in 2010. By the end of 2008, there was a total of 225 million rural-urban migrants.¹

Three stylized facts of this rural-urban migration emerged in recent years. First, shorter-distance migration is much more common than longer-distance migration. For example, migrants in coastal cities mostly come from rural areas in local or nearby provinces. Relatively few rural people in the West or North migrate to coastal provinces in the East and South, although they have much more to gain economically from such long-distance migration. Poncet (2006) documents that migration flows decrease significantly with the distance between origin and destination locations; intra-province migration flows are higher than inter-province flows and migration to adjacent provinces is more common than migration to provinces further away.² Our own survey data on rural-urban migrants in 15 cities show that about half of them come from rural areas within the local province.

Second, the earnings of these migrants vary substantially, depending on where they have migrated. Table 1 shows the average monthly earnings for rural-urban migrant household heads in the top 15 destination cities. This average varies widely across cities. On the top is Shanghai, where the average migrant makes 2,338 yuan a month. At the bottom is Chongqing, where the average is only 1,297 yuan, 45 percent lower. One might wonder whether these variations simply reflect different characteristics of migrants in different cities. The right column of Table 1 reports regression adjusted monthly earnings, controlling for gender, age, education, and experience in urban sectors. The variation pattern is the same:

¹These migrants hold a rural *hukou* but live and work in cities. They are generally referred to as *nong min gong*, meaning “farmers-turned workers” in Chinese.

²Some other studies such as Lin et al. (2004) and Bao et al. (2009), although not exactly focusing on the same question, have also noted a negative relationship between migration flow and distance.

Table 1: Average monthly earnings of migrant household heads in the top 15 migration destination Cities, 2008

City	Average monthly earnings (yuan)	Average monthly earnings (yuan), regression adjusted
Bengbu	1,778.31	1,761.68
Chengdu	1,751.30	1,685.26
Chongqing	1,296.64	1,300.19
Dongguan	1,445.46	1,430.70
Guangzhou	1,631.90	1,689.94
Hangzhou	2,254.95	2,246.65
Hefei	1,933.50	1,895.45
Luoyang	1,412.14	1,409.34
Nanjing	1,834.70	1,849.22
Ningbo	1,681.06	1,682.63
Shanghai	2,338.00	2,385.93
Shenzhen	1,859.85	1,818.25
Wuhan	1,551.69	1,528.91
Wuxi	1,748.05	1,824.82
Zhengzhou	1,396.08	1,394.77

Statistics in this table are based on a sample of 4,434 migrant household heads between 20 and 60 years old. The first column reports the simple average in each city. For the second column, we first regress monthly earnings on gender, age, years of schooling, urban experience (years since first migrated out of rural area), and city fixed effects, and then use the estimated coefficients to predict the average earnings in each city for the person with all independent variables set equal to sample means.

rural-urban migrants have very different income levels in different cities.

And third, due to an increased cost to attract migrant workers from far inland to coastal regions, there has emerged a trend that labor-intensive industries move from coastal to inland China to take advantage of the cheaper labor there. This trend has become so pervasive that many observers call it an “inward-moving wave.” A 2010 survey reveals that 21 percent of coastal manufacturers were considering relocating to inland regions.³ The most salient example is perhaps Foxconn, a contract manufacturer that hires more than 400,000 migrant workers in the coastal city Shenzhen and manufactures many renowned products such as iPod, iPad, and iPhone. In 2010, Foxconn announced the plan to construct new plants in inland cities such as Zhengzhou, Wuhan, and Chengdu; it would move the majority of its operations out of Shenzhen.

We argue that a simple phenomenon—migrants who grew up in rural China are reluctant to move far away from their birthplaces—helps explain all these three stylized facts. Partly because these migrants tend to avoid long-distance migration, we observe shorter-distance migration more often. It is for the same reason that migrant earnings are far from being equalized across cities; for cities with limited surplus labor in nearby rural areas, higher wages are necessary to attract migrant workers from remote regions. Originally, the labor intensive industries, especially those contract manufacturers, were highly concentrated in coastal regions, taking advantage of preferential policies in coastal economic development zones as well as the lower transportation costs for international trade. In recent years, the preferential policies have become ubiquitous and the transportation infrastructure in inland areas has improved substantially. As a result, the cost of hiring migrant workers has become a more prominent factor in firms’ locational decisions, which explains the “inward-moving wave” of labor-intensive industries.

There are many possible reasons as to why rural-urban migrants prefer shorter-distance moves. When an individual migrates to a city far from her birthplace, she will be disconnected from her social-family network, a most reliable source of emotional, physical, psychological, and sometimes even financial support in rural communities. She may have to live in an unfamiliar environment with different weather, food, and culture. She may feel isolated and insecure, and worry about being discriminated. For all of these reasons, one would be willing to give up some income in order to stay closer to home. Using recent survey data on a representative sample of 5,000 rural-urban migrant households in 15 cities, we empirically investigate this tradeoff between migration distance and expected income.

We build a simple model in which migrants from rural areas choose among a set of destination cities to maximize utility. The distance between a destination city and the individual’s home village is explicitly included in the utility function. We first estimate an individual’s expected income in each potential destination city using a semi-parametric method, controlling for potential self-selection biases. We then estimate the indirect utility

³See <http://finance.ifeng.com/roll/20100917/2631649.shtml> (viewed on February 19, 2011).

function for rural-urban migrants in China based on their migration patterns. We try different specifications including the conditional logit, nested logit, and mixed logit. We interact personal characteristics with migration distance and city characteristics to allow for heterogeneous preferences.

Our findings suggest that to induce an individual to migrate 10 percent further away from home, the wage paid to this migrant has to increase by 15 percent. This elasticity varies only slightly with distance; it is a little higher for female than male migrants; it is not affected by the migrant's age, education, or marital status. We discuss various policy implications of these findings.

The rest of the paper is organized as follows. Section 2 presents a simple model of migration destination choice. Section 3 describes the data we use and the construction of key variables. Section 4 presents empirical results. Section 5 concludes.

2 A Model of Migration Destination Choice

2.1 Basic setup

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 any of the J cities.⁴ If living in city j , individual i faces the following utility-maximization problem

$$\begin{aligned} \max U_{ij} &= C_{ij}^{\alpha_C} H_{ij}^{\alpha_H} D_{ij}^{-\beta} \exp [g(X_j) + \xi_j + \eta_{ij}] \\ \text{s.t. } & C_{ij} + \rho_j H_{ij} = I_{ij} \end{aligned} \quad (1)$$

- 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 non-tradable composite good (including, e.g., housing) in city j ; its price in city j is ρ_j .⁵

- D_{ij} is the distance from i 's home village to city j .

- X_j is a vector of characteristics (e.g., quality of air or public facilities) of city j ; g is a nonparametric function that we will not estimate here.

- 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.

- I_{ij} is i 's income in city j .

⁴In our empirical analysis, we will focus on household heads only, assuming that they are the decision makers.

⁵In addition to housing, many other goods can be considered as nontradable in China, which is especially true for rural-urban migrants because they do not have urban *hukou*. For example, depending on local regulations, rural-urban migrants may or may not have access to the heavily subsidized public schools and healthcare system in a city. So these migrant households face very different prices of education and healthcare in different cities.

Note that we include the migration distance in the utility function to capture the psychological costs associated with long-distance migration. We expect that migration distance causes disutility, thus the parameter β (with a minus sign in front) is expected to be positive.

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

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

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

$$\begin{aligned} U_{ij}^* &= \left(\frac{\alpha_C I_{ij}}{\alpha_C + \alpha_H} \right)^{\alpha_C} \left(\frac{\alpha_H}{\alpha_C + \alpha_H} \frac{I_{ij}}{\rho_j} \right)^{\alpha_H} D_{ij}^{-\beta} \exp [g(X_j) + \xi_j + \eta_{ij}] \\ &= \delta I_{ij}^{\alpha} D_{ij}^{-\beta} \exp [-\alpha_H \ln \rho_j + g(X_j) + \xi_j + \eta_{ij}]. \end{aligned}$$

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

$$V_{ij} = I_{ij}^{\alpha} D_{ij}^{-\beta} \exp [-\alpha_H \ln \rho_j + g(X_j) + \xi_j + \eta_{ij}]. \quad (2)$$

Denote WTP_i (i 's marginal *willingness to pay*) as the amount of money i is willing to give up in order to live closer to home village. From equation (2), this willingness to pay equals the marginal rate of substitution between migration distance and income, i.e.,

$$WTP_i = -\frac{\partial V_{ij} / \partial D_{ij}}{\partial V_{ij} / \partial I_{ij}} = \frac{\beta}{\alpha} \frac{I_{ij}}{D_{ij}}.$$

Taking the natural log of equation (2) and holding the utility level constant, we could also interpret $\frac{\beta}{\alpha}$ as an income-distance elasticity:

$$\frac{\beta}{\alpha} = \frac{\partial \ln I_{ij}}{\partial \ln D_{ij}} \approx \frac{\Delta I_{ij} / I_{ij}}{\Delta D_{ij} / D_{ij}}.$$

That is, to induce an individual to migrate 1 percent further away from home, one needs to offer this person an income that is $\frac{\beta}{\alpha}$ percent higher. Our goal in this paper is to empirically estimate α and β so that we can calculate this elasticity and the willingness to pay. To avoid cluttering notations, we treat β as a constant for the moment. Later we will allow β to vary with distance or individual characteristics in some of our empirical specifications.

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

$$\ln I_{ij} = \ln \hat{I}_{ij} + \varepsilon_{ij}. \quad (3)$$

We will estimate $\ln \hat{I}_{ij}$ based on individual i 's characteristics and the earnings of migrants

in city j , controlling for potential self-selection biases. This estimation procedure will be explained in detail in the next section on data and variables.

Following Timmins (2007), we assume that the price of the non-tradable good varies with city characteristics. For example, if a city has a fast growing-economy, low pollution, low congestion, and low crime rate, then one has to pay more for the non-tradable goods in order to live in the city.⁶ Specifically, we assume a flexible function

$$\ln \rho_j = h(X_j) + \epsilon_j \quad (4)$$

where h is a nonparametric function and ϵ_j an error term.

Substitute equations (3) and (4) into (2) and take natural logs to get

$$\ln V_{ij} = \alpha \ln \hat{I}_{ij} - \beta \ln D_{ij} + \theta_j + v_{ij} \quad (5)$$

where $\theta_j = g(X_j) - \alpha_H h(X_j) - \alpha_H \epsilon_j + \xi_j$ and $v_{ij} = \alpha \epsilon_{ij} + \eta_{ij}$. Note that everything in θ_j is fixed at the city level, so we may treat θ_j as a city fixed effect.

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

$$\Pr(\ln V_{ij} > \ln V_{ik} \forall k \neq j) = \frac{\exp(\alpha \ln \hat{I}_{ij} - \beta \ln D_{ij} + \theta_j)}{\sum_{s=1}^J \exp(\alpha \ln \hat{I}_{is} - \beta \ln D_{is} + \theta_s)}.$$

Therefore, the probability that every migrant i is living in city j as observed in the data is given by

$$L = \prod_i \prod_{j=1}^J \left[\frac{\exp(\alpha \ln \hat{I}_{ij} - \beta \ln D_{ij} + \theta_j)}{\sum_{s=1}^J \exp(\alpha \ln \hat{I}_{is} - \beta \ln D_{is} + \theta_s)} \right]^{\kappa_{ij}}, \quad (6)$$

where κ_{ij} is an indicator function that equals 1 if individual i is observed in city j . We will estimate $\{\alpha, \beta, \theta_1, \dots, \theta_J\}$ by maximizing this likelihood function.⁷ Note that if any set of parameters maximizes the likelihood function, then adding a constant to every θ_j will also maximize the likelihood function. That is, the absolute scales of $\{\theta_1, \dots, \theta_J\}$ are not identified. In practice, we will set $\theta_1 = 0$ (for the city of Guangzhou) and interpret each of the estimated θ_j as the difference from θ_1 .

Given our focus on α and β , we do not intend to estimate how observed city characteristics in X_j affect θ_j through functions g and h .⁸ In this baseline specification, we dump

⁶One can easily derive a relationship like this in a Rosen-Roback type model. See, e.g., Roback (1982).

⁷The conditional logit approach is commonly used for the analysis of migration choice. See, for example, Davies et al. (2001) and Poncet (2006), both of which use aggregate data for their empirical analysis. In contrast, we use individual level data to estimate the model here.

⁸Conceptually, function g determines how various city characteristics enter an individual's utility function; together with other parameters in the utility function, it determines how much this individual *is willing to pay* for the city characteristics. Function h , in contrast, shows how much an individual *has to pay* for these city characteristics. It reflects how much marginal local residents are willing to pay for the city characteristics (market demand for X) as well as the cost of maintaining such characteristics (supply of X).

the effects of both observed and unobserved city characteristics into the city fixed effect. In alternative specifications below, we will allow the preference for observed city characteristics X_j to vary across individuals and take the differential effects out of the city fixed effect.

2.2 Alternative specifications of the model

2.2.1 Nonconstant disutility of migration distance

The distaste for migration distance (β) is not necessarily a constant. We shall allow it to vary with distance or individual characteristics.

First, it is likely that the marginal disutility from long-distance migration will decline eventually. For example, if a migrant is only 100 km away from home village, then moving away for another 100 km may incur a substantial psychological cost. However, if the migrant is already 2,000 km away, another 100 km perhaps means very little. We explore this possibility by specifying β as a piecewise function:

$$\beta = \beta_1 1_{Q1} + \beta_2 1_{Q2} + \beta_3 1_{Q3} + \beta_4 1_{Q4} \quad (7)$$

where 1_{Qn} , $n \in \{1, 2, 3, 4\}$, is an indicator function that equals 1 if D_{ij} is in the n th quartile of the distribution of migration distance. Substituting this function for β in the likelihood function (equation (6)), we can estimate $\{\alpha, \beta_1, \beta_2, \beta_3, \beta_4, \theta_1, \dots, \theta_J\}$ through maximum likelihood.

Second, one might expect β (and thus WTP_i) to vary with individual characteristics such as gender, age, education, and marital status. To explore this possibility, we try an alternative specification in which β is assumed to vary across individuals and is determined in the following way:

$$\beta_i = b_0 + b_1 G_i + b_2 A_i + b_3 E_i + b_4 M_i \quad (8)$$

where G_i is individual i 's gender (=1 if male); A_i is i 's age; E_i is i 's years of schooling; and M_i indicates whether individual i is married. Again, substituting this function for β in the likelihood function (equation (6)), we can estimate $\{\alpha, b_0, b_1, b_2, b_3, b_4, \theta_1, \dots, \theta_J\}$ through maximum likelihood.

2.2.2 Differential preferences for observed city characteristics.

In addition to β , the preferences for observed city characteristics may also vary with individual characteristics. For example, younger migrants may have a stronger preference for larger cities because such cities offer a wider range of life opportunities. Similarly, better educated migrants may have a stronger preference for high-amenity cities. Although such differential preferences are not our focus in this study, we are concerned that uncontrolled heterogeneities may bias our estimates of the key parameters β and α . Thus we also experiment with alternative specifications to take into account differential preferences.

Specifically, we assume that individual i 's utility from K different characteristics of city j is

$$\Omega_{ij} = \bar{c}_j + \sum_{k=1}^K \left[c_{1k} \left(G_i X_j^k \right) + c_{2k} \left(A_i X_j^k \right) + c_{3k} \left(E_i X_j^k \right) + c_{4k} \left(M_i X_j^k \right) \right] \quad (9)$$

where G_i , A_i , E_i , and M_i are the same as defined above, X_j^k is city j 's characteristic k , and \bar{c}_j is the average utility derived from all such characteristics of city j . Notice that when we estimate the baseline model by maximizing the likelihood function given in equation (6), we essentially assume $c_{1k} = \dots = c_{4k} = 0$ and let \bar{c}_j be captured by the city fixed effect θ_j . Here we relax the first assumption but \bar{c}_j is still unidentifiable due to the inclusion of the city fixed effects. Therefore, we estimate the parameters by maximizing the following likelihood function

$$\tilde{L} = \prod_i \prod_{j=1}^J \left\{ \frac{\exp[\alpha \ln \hat{I}_{ij} - \tilde{\beta} \ln D_{ij} + (\Omega_{ij} - \bar{c}_j) + \theta_j]}{\sum_{s=1}^J \exp[\alpha \ln \hat{I}_{is} - \tilde{\beta} \ln D_{is} + (\Omega_{is} - \bar{c}_s) + \theta_s]} \right\}^{\kappa_{ij}}$$

where we substitute equation (9) for Ω_{ij} and may replace $\tilde{\beta}$ with the right-hand side of equation (7) or (8), depending whether and how we allow the parameter β to vary.

2.2.3 Nested logit

The conditional-logit setup in the specifications above assumes the *independence from irrelevant alternatives* (IIA).⁹ This might be violated given that some of the destination cities in our data are physically close to each other. Some of them are in the same province or in the same region. For example, Dongguan, Shenzhen, and Guangzhou all belong to the Guangdong province and are in the Pearl River Delta region. Thus we try the nested logit as an alternative specification.

Allowing for differential preferences, we rewrite the log indirect utility as

$$\begin{aligned} \ln V_{ij} &= \alpha \ln \hat{I}_{ij} - \beta \ln D_{ij} + \sum_{k=1}^K \left[c_{1k} \left(G_i X_j^k \right) + c_{2k} \left(A_i X_j^k \right) + c_{3k} \left(E_i X_j^k \right) + c_{4k} \left(M_i X_j^k \right) \right] \\ &\quad + \sum_{s=1}^J \theta_s \kappa_{ij} + v_{ij} \\ &= \Psi_{ij} \Upsilon + v_{ij} \end{aligned} \quad (10)$$

where

$$\Psi_{ij} = \left(\ln \hat{I}_{ij}, -\ln D_{ij}, G_i X_j^1, A_i X_j^1, E_i X_j^1, M_i X_j^1, \dots, G_i X_j^K, A_i X_j^K, E_i X_j^K, M_i X_j^K, \kappa_{i1}, \dots, \kappa_{iJ} \right)$$

⁹Let P_{ij} be the probability of individual i choosing city j . IIA means that P_{ij}/P_{ik} is independent of the characteristics (and even the existence) of any city other than j and k .

and $\Upsilon = (\alpha, \beta, c_{11}, c_{21}, c_{31}, c_{41}, \dots, c_{1K}, c_{2K}, c_{3K}, c_{4K}, \theta_1, \dots, \theta_J)$.

Let N be the number of destination regions (“nests”) and B_n the set of destination cities in region n . Following McFadden (1978), we now assume that v_{ij} follows a generalized extreme value (GEV) distribution with the cumulative density function

$$F = \exp \left[- \sum_{n=1}^N \left(\sum_{j \in B_n} e^{-v_{ij}/\lambda_n} \right)^{\lambda_n} \right]$$

where the parameter λ_n is a measure of the degree of independence in unobserved utility among the alternatives in nest n .¹⁰ Then for any $j \in B_n$, the probability of i choosing j is

$$\Pr(i \text{ in } j \in B_n) = \frac{\exp(\Psi_{ij}\Upsilon/\lambda_n) [\sum_{s \in B_n} \exp(\Psi_{is}\Upsilon/\lambda_n)]^{\lambda_n - 1}}{\sum_{m=1}^N [\sum_{q \in B_m} \exp(\Psi_{iq}\Upsilon/\lambda_m)]^{\lambda_m}}.$$

Therefore, Υ can be estimated through maximizing the likelihood function

$$\ddot{L} = \prod_i \prod_{j=1}^J \prod_{n=1}^N \left\{ \frac{\exp(\Psi_{ij}\Upsilon/\lambda_n) [\sum_{s \in B_n} \exp(\Psi_{is}\Upsilon/\lambda_n)]^{\lambda_n - 1}}{\sum_{m=1}^N [\sum_{q \in B_m} \exp(\Psi_{iq}\Upsilon/\lambda_m)]^{\lambda_m}} \right\}^{\kappa_{ijn}}.$$

The indicator function κ_{ijn} takes value one if i chooses city j and j is in region n , and zero otherwise.

2.2.4 Mixed logit

Although we allow β to vary, we have imposed stringent functional-form restrictions on how it varies. In this alternative specification, we treat the two key parameters, β and α , as random variables across individuals. We assume that each follows a distribution but impose nothing on how it varies across individuals. We estimate the distributions of β and α based on a mixed logit model.¹¹ We then use their mean values to calculate *WTP* and the income distance elasticity.

We again specify the indirect utility function as in equation (10), allowing for heterogeneous preferences for all city characteristics:

$$\ln V_{ij} = \Psi_{ij} \tilde{\Upsilon} + v_{ij}. \quad (11)$$

The tilde on top of Υ indicates that some coefficients are now random.

We assume:

- (i) v_{ij} follows an i.i.d. type I extreme value distribution; and

¹⁰As is well known, this nested logit model reduces to the standard logit model if $\lambda_n = 1 \forall n$ (McFadden, 1978).

¹¹The mixed logit model (*aka* random-coefficients logit) actually allows us to treat any set of parameters in the utility function as random across individuals. However, assuming random preferences for other city characteristics will necessarily change the city fixed effects specification. More specifically, because city characteristics are all unique to each city, one has to drop some city fixed effects in order to add those city characteristics; otherwise, there will be perfect collinearity.

(ii) $\tilde{\Upsilon}$ has a density function $f(\tilde{\Upsilon}|\Lambda)$, where Λ represents the parameters of this distribution such as the mean and covariance of $\tilde{\Upsilon}$.¹²

Then the probability of i choosing j is

$$\Pr(i \text{ in } j) = \int \frac{\exp(\Psi_{ij}\tilde{\Upsilon})}{\sum_{s=1}^J \exp(\Psi_{is}\tilde{\Upsilon})} f(\tilde{\Upsilon}|\Lambda) d\tilde{\Upsilon}.$$

Following standard practice, we will assume that the density f is normal or log-normal. Given the high dimensionality of $\tilde{\Upsilon}$, this probability generally cannot be solved analytically. We thus approximate it through simulation (McFadden and Train, 2000; Train, 2009, ch. 6).

Given any value Λ , we will (i) randomly draw a value from $f(\tilde{\Upsilon}|\Lambda)$ and label it $\tilde{\Upsilon}^r$ with the superscript indicating this as the r th draw; and (ii) evaluate the logit formula $\frac{\exp(\Psi_{ij}\tilde{\Upsilon})}{\sum_{s=1}^J \exp(\Psi_{is}\tilde{\Upsilon})}$ with this draw. Repeat (i) and (ii) R times and calculate the average

$$\widehat{\Pr}(i \text{ in } j) = \frac{1}{R} \sum_{r=1}^R \frac{\exp(\Psi_{ij}\tilde{\Upsilon}^r)}{\sum_{s=1}^J \exp(\Psi_{is}\tilde{\Upsilon}^r)},$$

which is an unbiased estimator of the choice probability. A simulated log likelihood is then defined as

$$SLL = \sum_i \sum_{j=1}^J \kappa_{ij} \left[\frac{1}{R} \sum_{r=1}^R \frac{\exp(\Psi_{ij}\tilde{\Upsilon}^r)}{\sum_{s=1}^J \exp(\Psi_{is}\tilde{\Upsilon}^r)} \right],$$

where, again, $\kappa_{ij} = 1$ if i chooses j and zero otherwise.

The value of Λ that maximizes this SLL is called a maximum simulated likelihood estimator (MSLE). The estimate of Λ is then used to describe the distribution of $\tilde{\Upsilon}$. We need mean $\tilde{\alpha}$ and $\tilde{\beta}$ to calculate WTP and the income distance elasticity.

3 Data and Key Variables

For empirical analysis, we use a unique survey database on Rural-Urban Migration in China (RUMiC). As part of a large research project, the database is being constructed by a team of researchers from Australia, China, and Indonesia. They secured funding from various sources to conduct a five-year longitudinal survey in China and Indonesia, with the goal of studying

¹²We may write $\tilde{\Upsilon}$ as the sum of its mean and a random deviation: $\tilde{\Upsilon} = \Upsilon + \sigma_{\Upsilon}$. Then the random-coefficient indirect utility (equation 10) is $\ln V_{ij} = \Psi_{ij}\Upsilon + (\Psi_{ij}\sigma_{\Upsilon} + v_{ij})$. Note that the first term still has constant coefficients Υ . We may consider the whole second part $(\Psi_{ij}\sigma_{\Upsilon} + v_{ij})$ as the stochastic component of the utility. Thus we can also derive the random-coefficient model by imposing conditions on the error term of a constant-coefficient model. More specifically, consider the indirect utility function $\ln V_{ij} = \Psi_{ij}\Upsilon + \mu_{ij}$, where Υ is constant. Let us assume the error term has two components: $\mu_{ij} = \Psi_{ij}\sigma_{\Upsilon} + v_{ij}$. The first part is random, governed by a density function $f(\tilde{\Upsilon}|\Lambda)$, and the second part follows an i.i.d. type I extreme value distribution. Then we have a model exactly the same as the random-coefficient logit. Indeed, it is well-known that the random-coefficient and error-component specifications of the mixed logit model are equivalent (Train, 2009, ch. 6). From the error-component interpretation, we immediately recognize that this mixed logit does not require the IIA assumed by the standard logit model. In fact, mixed logit can approximate any substitution pattern among alternatives (McFadden and Train, 2000).

issues such as the effect of rural-urban migration on income mobility and poverty alleviation, the state of education and health of children in migrant families, and the assimilation of migrant workers into the city.

We use the first wave of the survey data, for which the survey was conducted in 2008 and the data became available in 2009. In China, three representative samples of households were surveyed, including a sample of 8,000 rural households, a sample of 5,000 rural-urban migrant households, and a sample of 5,000 urban households. In this paper, our empirical analyses use information mainly from the migrant sample. Since the migrants all came from rural areas, 99.4 percent of them have a rural *hukou*, although they currently live in cities.

The migrants surveyed were randomly chosen from 15 cities that are the top rural-urban migration destinations in China.¹³ Eight of these cities are in coastal regions (Shanghai, Nanjing, Wuxi, Hangzhou, Ningbo, Guangzhou, Shenzhen, and Dongguan); five of them are in central inland regions (Zhengzhou, Luoyang, Hefei, Bengbu, and Wuhan); and two of them are in the west (Chengdu and Chongqing).

Figure 1 shows a map of China and highlights the 15 cities where the migrant survey was conducted. It is important to note that these cities are scattered over different regions in China. This implies that for a typical migrant in our database, the migration distance to different destinations varies substantially. This large variation in migration distance, together with the already mentioned variation in monthly earnings across cities, is crucial for us to precisely estimate the income distance tradeoff.

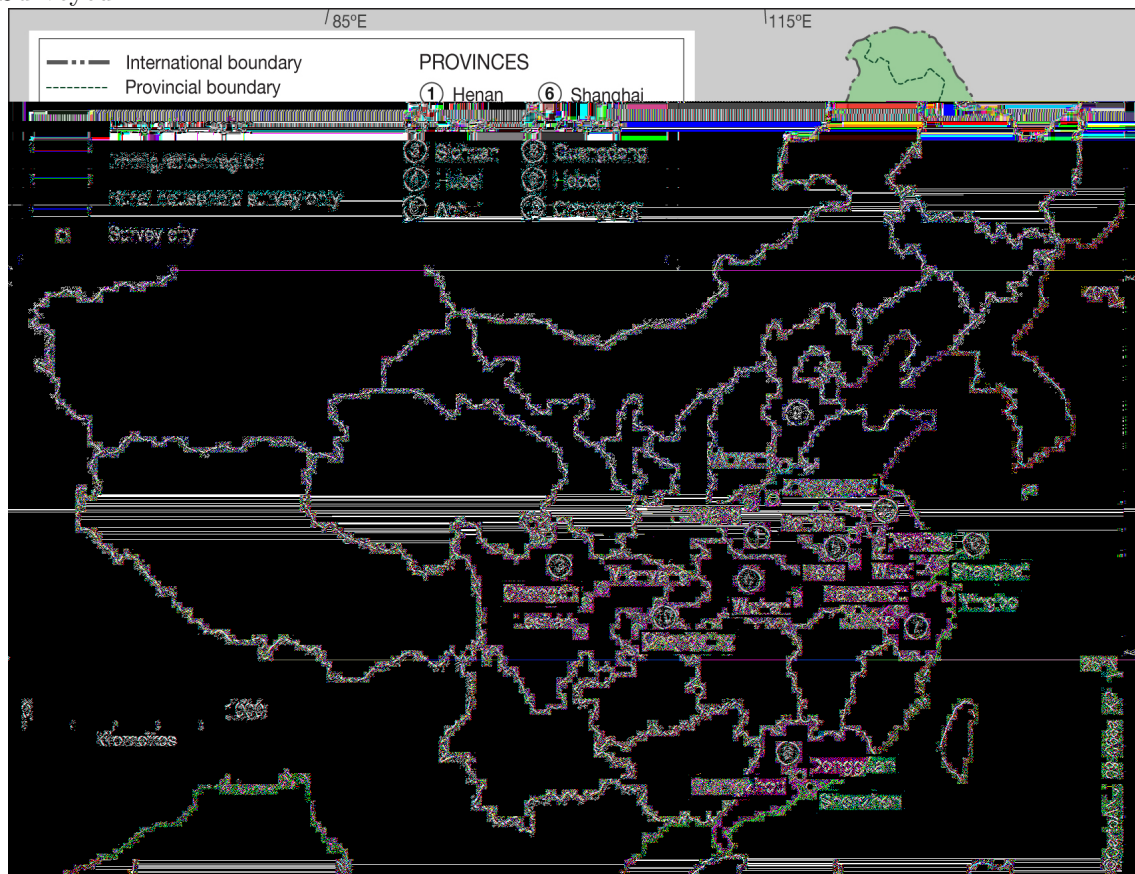
Although our analysis in this paper focuses on household heads, the migrant survey actually collected information about every household member. It asked detailed questions about the respondent's personal characteristics, educational background, employment situation, health status, children's education, social and family relationship, major life events, income and expenditure, housing and living conditions, etc. The resultant database contains more than 700 variables. In terms of basic information of a household member, we know the person's age, gender, education level, current address, home address before migration, etc. For information regarding employment experience, we know whether the person is self-employed or a wage worker, occupation, monthly income, how he/she found the current job, what was his/her first job, how he/she found the first job, etc.

Before implementing the maximum likelihood estimation, we need to calculate the distance from each individual i 's home village to every city j (D_{ij}). We also need the predicted income for each individual i in each city j ($\ln \hat{I}_{ij}$), which is not directly observed in the data.

For every migrant household head, the survey has asked about his or her home address. This field of information is recorded in Chinese, which appears to have many errors because the character-based language has different intonations and is prone to spelling errors. We first clean the home address information down to the home county level. Using an on-

¹³A sampling procedure was very carefully designed to ensure that migrants in the database constituted a representative sample of all the migrants in the 15 cities. See the RUMiCI Project's homepage (<http://rumici.anu.edu.au/joomla/>) for detailed documentation of the sampling method.

Figure 1: The Top Fifteen Destination Cities in China Where Rural-Urban Migrants Were Surveyed



Source: The Rural-Urban Migration in China and Indonesia Project Website (http://rumici.anu.edu.au/joomla/index.php?option=com_content&task=view&id=49&Itemid=52), with modifications. The rural-urban migrants are surveyed in the 15 cities that are highlighted with blue rectangles. Urban households are surveyed in all the 18 cities on this map.

line data source, we find the latitude-longitude coordinate for each home county and each destination city.¹⁴ We then use the Haversine formula to calculate the “great-circle distance” (on the surface of earth) from the home county to each city.¹⁵

In theory, physical, cultural, and social distances perhaps all matter in one’s migration decision. Here we use the physical distance only and assume that other relevant distances are highly correlated with physical distance. Even for physical distance, one might argue that railway or highway distance is more relevant. However, such information at the county level is difficult to obtain and changes almost daily because China has been continuously upgrading its transportation infrastructure. We therefore use the “great-circle distance” as a proxy.

To generate $\ln \hat{I}_{ij}$, we run a series of city-specific regressions of income on individual characteristics. We use these estimates to predict $\ln \hat{I}_{ij}$. A simple OLS regression for each city is likely to produce biased estimates because of sorting across cities. We follow a semi-parametric approach to correct the potential selection biases. The methodology is developed by Dahl (2002) and used by Bayer et al. (2009).¹⁶

Consider the following model

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

where $\ln I_{ij}$ is log income for individual i in city j ; Z_i is a vector of individual characteristics; and μ_{ij} is 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 .

¹⁴The online data source is <http://ngcc.sbsm.gov.cn/Mapquery/default.aspx>, the website of the National Geomatics Center of China.

¹⁵Let $(lat_j, long_j)$ and $(lat_k, long_k)$ be the latitude-longitude coordinates of two locations j and k . Then the shortest distance between j and k over the earth’s surface, d , can be calculated using the Haversine formula (Sinnott, 1984):

$$\begin{aligned} \Delta lat &= lat_k - lat_j \\ \Delta long &= long_k - long_j \\ a &= \left[\sin \left(\frac{\Delta lat}{2} \right) \right]^2 + \cos(lat_j) \cdot \cos(lat_k) \cdot \left[\sin \left(\frac{\Delta long}{2} \right) \right]^2 \\ c &= 2 \cdot \text{atan2}(\sqrt{a}, \sqrt{1-a}) \\ d &= R \cdot c \end{aligned}$$

where R is the earth’s radius (with a mean value of 6,371 km). Note that angles need to be in radians.

¹⁶It has long been recognized that there is a problem of self-selection when estimating income for migrants. 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), because Monte Carlo simulations suggest that Dahl’s method is preferred to Lee’s (Bourguignon et al., 2007).

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

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

where P_{ij} is the probability of i choosing j and $\psi(\cdot)$ is an unknown function that gives the conditional mean $E(\mu_{ik}|\cdot)$. Dahl (2002) introduces an “index sufficiency assumption,” assuming that the probability of the first-best choice is the only information needed for the estimation of the conditional mean. This dramatically reduces the dimension of the correction function ψ and the above estimation equation becomes

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

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 polynomial or Fourier series expansions.

Following this approach, for each destination city j , we use the information about all the individuals who migrated to this city to estimate an equation for log income. Our goal is to predict each migrant’s income in city j , regardless where she actually migrated.

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 province and education level. We identify the top eight home provinces in our data and lump the rest of the provinces into an “other home provinces” category.¹⁷ Within each of the nine home-province groups, individuals are further 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). Thus we have put all the individuals into 18 different cells.¹⁸ For each individual i in city j , we find the cell she belongs to. The estimated probability of i choosing j , \hat{P}_{ij} , is simply calculated as the proportion 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} :

$$\ln I_{ij} = Z_i \gamma_j + b_{j1} \hat{P}_{ij} + b_{j2} (\hat{P}_{ij})^2 + \varepsilon_{ij}.$$

Included in Z_i are age, age squared, gender, years of schooling, marital status, self-employment

¹⁷It is not entirely arbitrary to choose the cutoff at the eighth largest home province. These eight provinces actually cover all of the destination cities except Shanghai. Shanghai itself is a province-level jurisdiction. However, only three migrants come from rural areas in Shanghai. The group is too small to be treated as a separate one.

¹⁸There is a tradeoff between having more cells and the precision of estimated migration probability. Because each individual can choose among 15 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.

status, and a constant.¹⁹ This regression only uses the information on migrants in city j . We then use $\hat{\gamma}_j$ to predict $\ln \hat{I}_{mj}$ for every individual m in our sample. Note that we add \hat{P}_{ij} and its square term to the regression only for estimating an unbiased $\hat{\gamma}_j$; we do not need them when predicting income.

Finally, we have also collected information on destination city characteristics from the *Urban Statistical Yearbook of China*.²⁰ We construct nine variables at the city level, including population size, per capita GDP, five-year average unemployment rate, per capita elementary schools, per capita hospital beds, per capita public buses, per capita paved road area, per capita green area (lawn, flower beds, etc.), and per capita air pollutants emitted by industries. We will include these variables in some of our empirical specifications to allow for differential preferences for observed city characteristics.

4 Empirical Results

We present empirical results in this section.

4.1 Descriptive statistics

Our analysis uses the data on 5,000 rural-urban migrant households in China. We focus on the household heads only. Dropping those younger than 20 and older than 60, we end up with 4,434 migrants, for which we present some descriptive statistics in Table 2.

Seventy-one percent of these migrants are male; 61 percent of them are married. The average person is 32 years old, has 9.3 years of education, and makes 1,759 yuan a month. The average log migration distance is 5.364; this distance has a wide range from 1.557 (4.75 km) to 8.309 (4,061 km). Fifty-five percent of these migrants are from the local province.

The average migrant first moved to a city 8.5 years ago. Note that this does not mean that the person has lived and worked in the city for all these years. There might be some time in between when the migrant returned to the home village for some reason and then migrated out again later. It is also important to note that migrants do not necessarily settle down after migration. Indeed, a quarter of the migrants in the sample are currently not in their first destination provinces. That is, a migrant might first migrate to province A, but later found a better job opportunity in province B and thus moved to B. Similarly, many of these migrants also moved from one job to another; 60 percent of them are currently not in their first jobs in urban sectors. This indicates that migrants indeed reoptimize as new information or opportunities come up over time, which is important because we model them as utility maximizers.

¹⁹A polynomial function also has a constant, but we cannot include it in the regression because of this constant in Z_i . It is impossible to separately identify both of them.

²⁰We mainly use the 2008 edition of the yearbook, which reports information from 2007. For unemployment rates, we also use four earlier editions so that we can calculate a five-year average.

Table 2: Descriptive statistics for migrant household heads

	Mean	Std. Dev.	Minimum	Maximum
Male	0.709	0.454	0	1
Age	31.80	9.46	20	60
Years of schooling	9.26	2.45	1	20
Married	0.605	0.489	0	1
Monthly earnings	1,758.67	2,508.09	0	99,998
Log migration distance	5.364	1.153	1.557	8.309
From local province	0.554	0.497	0	1
Years since first migrated out of village	8.49	6.47	0	45
Still in first destination province	0.747	0.435	0	1
Still in first job in urban sectors	0.398	0.490	0	1

Statistics in this table are based on a sample of 4,434 migrant household heads between 20 and 60 years old.

4.2 Regression results

We start with the baseline specification that only includes log income, log distance, and city fixed effects. The results are in column (1) of Table (3). The coefficient is 1.05 for log income and 2.09 for (negative) log distance, both are statistically significant at very high levels of confidence. The estimated $\frac{\beta}{\alpha}$ is 1.99, which is also highly significant. This estimate implies that income has to increase by 20 percent to induce the average migrant to move 10 percent further away from home, which seems to be very high.

Although our focus is not on the city fixed effects, it is important to check whether their values make sense. Our reference city is Guangzhou, the third largest city in China and the main manufacturing hub in southern China. All city fixed effects are negative; they are all statistically significant except for Shanghai and Shenzhen. That is, if not for income and distance reasons, most other cities are less attractive to migrants than Guangzhou. The difference is the largest for Bengbu, Luoyang, and Chongqing, all inland cities in less developed regions. All of these seem to make sense. We examine the simple correlation between the city fixed effects and city characteristics. We find that the fixed effects are positively correlated with population size, per capita GDP, per capita elementary schools, per capita hospital beds, per capita public buses, per capita paved road area, and per capita green area, and that they are negatively correlated with five-year average unemployment rate and per capita air pollutants emitted. These are all exactly as expected.

In column (2) of Table (3), we estimate different values of β for different quartiles of migration distance. They are more or less the same, ranging from 2.06 to 2.12. Because these parameters are so precisely estimated, it turns out that 2.06 is statistically significantly different from 2.12. However, the size of the difference is so small that it has little economic significance. At the bottom of column (2), we also report the estimated $\frac{\beta}{\alpha}$ for different

Table 3: Regression results

Variable coefficient name	(1) Conditional Logit	(2) Conditional Logit	(3) Conditional Logit
Log income α	1.050 (0.151)	1.055 (0.151)	1.093 (0.154)
-Log distance β or b_0	2.091 (0.034)		2.093 (0.156)
-Log distance* 1_{Q1} β_1		2.089 (0.046)	
-Log distance* 1_{Q2} β_2		2.116 (0.041)	
-Log distance* 1_{Q3} β_3		2.102 (0.039)	
-Log distance* 1_{Q4} β_4		2.062 (0.038)	
-Log distance*male b_1			-0.122 (0.055)
-Log distance*age b_2			0.005 (0.003)
-Log distance*education b_3			-0.009 (0.010)
-Log distance*married b_4			0.046 (0.063)
City fixed effects	Yes	Yes	Yes
Log likelihood	-6,699.71	-6,689.46	-6,693.60
Number of observations	59,820	59,820	59,820
Post-regression estimation of $\frac{\beta}{\alpha}$	1.992 (0.288)	1.979 (0.286) 2.005 (0.289) 1.992 (0.287) 1.954 (0.281)	Female: 1.991 (0.280) Male: 1.879 (0.264)

Standard errors are in parentheses. There are 4,434 migrant household heads between 20 and 60 years old, but 446 of which are not used in these regressions due to missing variables. The number of observations equals the number of migrants (3,988) multiplied by the number of destination cities (15). For specification (2), $\frac{\beta}{\alpha}$ is calculated separately for the four different quartiles of migration distance.

quartiles. They are all close to 2. Therefore, it appears that the income-distance elasticity changes very little with distance, which is somewhat surprising.

In column (3) of Table (3), we allow β to vary with individual characteristics by adding the interactions between log distance and individual characteristics. Only being male is associated with a significantly lower β . Other individual characteristics, including age, education, and marital status, do not affect the coefficient of log distance. The estimated $\frac{\beta}{\alpha}$ is 1.99 for female migrants, in contrast to 1.88 for male migrants. In other words, it is relatively easier to induce male migrants to move further away from home than female migrants.

In Table (4), we present results from three specifications parallel to those in Table (3); the only difference is that now we allow for differential preferences over all observed city characteristics. More specifically, we add interactions between individual and city characteristics into the regression. We have four individual characteristics including gender, age, education, and marital status; we have nine city characteristics including population size, per capita GDP, five-year average unemployment rate, per capita elementary schools, per capita hospital beds, per capita public buses, per capita paved road area, per capita green area, and per capita air pollutants emitted. In total, there are 36 interaction terms added to the regression.

Comparing the results in Table (4) and (3), we see that the biggest difference is the coefficient of log income. It is now much higher: 1.4 as opposed to the earlier estimates that are all below 1.1. The coefficient of log distance is still close to 2. Therefore, the estimated $\frac{\beta}{\alpha}$ is lower now at about 1.5. That is, to induce a migrant to move 10 percent further away from home, the income needs to increase by 15 percent. Similar to the results in Table (3), this elasticity varies only slightly across different quartiles of migration distance, ranging from 1.48 to 1.52. We again find a significant difference between male and female migrants: whereas this elasticity is 1.57 for females, it is 1.45 for males.

We have again examined the simple correlation between the city fixed effects and city characteristics. Same as before, migrants appear to like cities with larger population, higher GDP, and better infrastructure and public facilities; they dislike cities with higher unemployment rates or severe air pollution. Although not presented in Table (4), some of the results regarding the interaction terms are interesting to note. For example, female migrants like larger cities, greener cities, and cities with lower air pollution more than male migrants; male migrants prefer cities with more paved roads and more public buses more than female migrants. Compared to less educated migrants, more educated ones dislike unemployment more and care less about per capita GDP or elementary schools. Older migrants also care less about elementary schools, perhaps because they do not have school-aged children any more.

In Table (5), we present results from nested logit regressions. In China, Pearl River Delta and Yangtze River Delta are the two leading commercial and manufacturing regions;

Table 4: Regression results

Variable coefficient name	(1) Conditional Logit	(2) Conditional Logit	(3) Conditional Logit
Log income α	1.391 (0.176)	1.396 (0.176)	1.414 (0.177)
-Log distance β or b_0	2.101 (0.035)		2.123 (0.209)
-Log distance* 1_{Q1} β_1		2.087 (0.047)	
-Log distance* 1_{Q2} β_2		2.117 (0.042)	
-Log distance* 1_{Q3} β_3		2.103 (0.039)	
-Log distance* 1_{Q4} β_4		2.068 (0.038)	
-Log distance*male b_1			-0.165 (0.075)
-Log distance*age b_2			0.002 (0.005)
-Log distance*education b_3			0.002 (0.014)
-Log distance*married b_4			0.020 (0.086)
Differential preferences for city characteristics	Yes	Yes	Yes
City fixed effects	Yes	Yes	Yes
Log likelihood	-6,578.29	-6,569.34	-6,575.70
Number of observations	59,820	59,820	59,820
Post-regression estimation of $\frac{\beta}{\alpha}$	1.511 (0.192)	1.494 (0.190) 1.516 (0.192) 1.506 (0.191) 1.481 (0.187)	Female: 1.571 (0.196) Male: 1.454 (0.182)

Standard errors are in parentheses. In each specification, the interactions of individual and city characteristics are included to allow for differential preferences. There are four individual characteristics (gender, age, education, and marital status), nine city characteristics (population, per capita GDP, 5-year unemployment rate, per capita elementary schools, per capita hospital beds, per capita public buses, per capita paved road area, per capita green area, per capita air pollutants), and therefore a total of 36 interactions. There are 4,434 migrant household heads between 20 and 60 years old, but 446 of which are not used in these regressions due to missing variables. The number of observations equals the number of migrants (3,988) multiplied by the number of destination cities (15). For specification (2), $\frac{\beta}{\alpha}$ is calculated separately for the four different quartiles of migration distance.

Table 5: Regression results

Variable coefficient name	(1) Nested Logit	(2) Nested Logit	(3) Nested Logit
Log income α	1.271 (0.179)	1.319 (0.186)	1.295 (0.179)
-Log distance β or b_0	1.989 (0.054)		1.963 (0.207)
-Log distance*1 $_{Q1}$ β_1		2.006 (0.071)	
-Log distance*1 $_{Q2}$ β_2		2.043 (0.068)	
-Log distance*1 $_{Q3}$ β_3		2.020 (0.064)	
-Log distance*1 $_{Q4}$ β_4		1.995 (0.061)	
-Log distance*male b_1			-0.181 (0.074)
-Log distance*age b_2			0.005 (0.005)
-Log distance*education b_3			0.004 (0.014)
-Log distance*married b_4			-0.045 (0.084)
Differential preferences for city characteristics	Yes	Yes	Yes
City fixed effects	Yes	Yes	Yes
Log likelihood	-6,551.78	-6,542.64	-6,548.39
p-value of LR test for IIA ($\lambda = 1$)	0.000	0.000	0.000
Number of observations	59,820	59,820	59,820
Post-regression estimation of $\frac{\beta}{\alpha}$	1.565 (0.214)	1.520 (0.207) 1.547 (0.210) 1.530 (0.208) 1.511 (0.206)	Female: 1.633 (0.226) Male: 1.493 (0.207)

Standard errors are in parentheses. We specify the nested logit model with three nests: (1) the Pearl River Delta region, including Guangzhou, Dongguan, and Shenzhen; (2) the Yangtze River Delta region, including Shanghai, Nanjing, Wuxi, Hangzhou, and Ningbo; (3) the rest of the country, including Zhengzhou, Luoyang, Hefei, Bengbu, Wuhan, Chongqing, and Chengdu. In each specification, 36 interactions of individual and city characteristics are included to allow for differential preferences (see the notes under Table (4) for more detailed explanation). There are 4,434 migrant household heads between 20 and 60 years old, but 446 of which are not used in these regressions due to missing variables. The number of observations equals the number of migrants (3,988) multiplied by the number of destination cities (15). For specification (2), $\frac{\beta}{\alpha}$ is calculated separately for the four different quartiles of migration distance.

they have their distinctive identities because of their economic prosperity in the post-reform era. For this reason, we lump all the cities in the Pearl River Delta region into one group (including Guangzhou, Dongguan, and Shenzhen), cities in the Yangtze River Delta region into the second group (including Shanghai, Nanjing, Wuxi, Hangzhou, and Ningbo), and all other cities into the third group. We are assuming that migrants first decide whether to migrate to the Pearl River Delta region, the Yangtze River Delta region, or the rest of the country; they will then choose a destination city among those within a region. We again allow the distance parameter to vary with migration distance or individual characteristics in two separate specifications. In all regressions, we include city fixed effects and control for differential preferences over observed city characteristics.

For all three nested-logit specifications, we test for IIA. In each case, it is rejected. That is, the IIA assumption in the conditional logit regressions is very unlikely to hold. However, the alternative nested logit specification has very limited effects on our key estimates. The estimated $\frac{\beta}{\alpha}$ is still close to 1.5. It does not vary much across different distance quartiles. Gender of the migrants still makes a difference: Whereas the ratio is 1.63 for females, it is 1.49 for male migrants. Therefore, although these nested logit models seem to be more reasonable than conditional logit models, they do not change any of our major findings.

Finally, in Table (6), we report regression results from mixed logit models. The two key parameters, α and β , are assumed to be independently normal in column (1) and independently log normal in column (2). The log normal assumption perhaps makes more sense because we expect both α and β to be positive. Under both specifications, we assume that other parameters are fixed. The estimated mean values of α and β are similar from these two specifications; they are slightly larger under the log normal specification. The estimated ratio $\frac{\beta}{\alpha}$ (based on their mean values) is close to 1.5 in both cases, which is similar to what we obtained from conditional and nested logit models.

Overall, we find our results are robust to alternative specifications. As long as we allow for differential preferences for observed city characteristics, the estimated $\frac{\beta}{\alpha}$ is always close to 1.5. Results from several specifications indicate that this elasticity is lower for male than female migrants.

To give these results some concrete meaning, we do the following exercise. Let's assume that we want to induce every migrant to move 10 percent further away from home, which for the average migrant is 38 km further away. This requires a 15.71 percent increase in earnings for a female migrant or a 14.54 percent increase for a male migrant (based on results in column (3) of Table (4)). In monetary terms, it means that the monthly earnings for the average migrant need to increase by 261 yuan. For labor intensive industries, there is clearly a great deal to gain by moving closer to rural regions with a large amount of surplus labor.

[***More back-of-envelope calculations, e.g., how much labor cost can Foxconn expect to save if they relocate all their plants from Shenzhen to Chengdu, Wuhan, and Zhengzhou?***]

Table 6: Regression results

Variable coefficient name	(1) Mixed Logit Independently normal α, β	(2) Mixed Logit Independently log normal α, β
Log income α	Mean: 1.620 (0.216) Standard deviation: 1.201 (0.412)	Mean: 1.724 (0.222) Standard deviation: 1.920 (0.495)
-Log distance β	Mean: 2.484 (0.060) Standard deviation: 0.919 (0.070)	Mean: 2.580 (0.079) Standard deviation: 1.162 (0.130)
Differential preferences for city characteristics	Yes	Yes
City fixed effects	Yes	Yes
Log likelihood	-6,520.13	-6,522.63
Number of observations	59,820	59,820
Post-regression estimation of $\frac{\beta}{\alpha}$	1.533 (0.202)	1.497 (0.192)

Standard errors are in parentheses. In each specification, 36 interactions of individual and city characteristics are included to allow for differential preferences (see the notes under Table (4) for more detailed explanation). There are 4,434 migrant household heads between 20 and 60 years old, but 446 of which are not used in these regressions due to missing variables. The number of observations equals the number of migrants (3,988) multiplied by the number of destination cities (15). These mixed logit models are estimated using the Stata module MIXLOGIT. It is created by the economist Arne Risa Hole (available at <http://www.sheffield.ac.uk/economics/people/hole/stata.html>).

4.3 Further discussion

One wonders why rural-urban migrants in China are willing to forgo so much income in order to stay closer to home.

The very first question one might ask is whether this willingness to pay reflects a higher pecuniary cost associated with long-distance moves. Note that in our theoretical model, we have completely ignored any monetary moving expenses, so part or all of such expenses might be captured by the distance coefficient. However, upon closer examination, we find that moving expenses in China are simply too low to be able to explain even a small part of these migrants' willingness to pay. A concrete example helps put this into perspective. Consider a trip from Wuhan to Guangzhou by express train. The total distance is 1,069 km and the ticket price for a "hard seat" is only 140 yuan.²¹ That is, on average it only costs 4.98 yuan to travel 38 km, which is close to zero compared to the 261 yuan a month the average migrant is willing to pay.

A second potential explanation is the lack of information about job openings in faraway cities. This could explain why long-distance migration flows are smaller than short-distance ones; however, it itself does not explain why long-distance moves are generally associated with higher earnings as shown in the data. Moreover, if the the lack of information is the reason for a high income-distance elasticity, then firms should have incentive to advertise jobs in regions with surplus labor instead of moving to those regions. And last, as argued by Schwartz (1973), if the negative effect of distance on migration destination choices is really an information effect, then we would expect it to decrease (in absolute value) with education. However, we find that education does not matter, which also suggests that the lack of information in remote regions does not explain this income-distance tradeoff.

Yet another possible explanation is that the willingness to pay for staying closer to home is really a willingness to pay for larger social networks. It is quite possible that in a city closer to one's home, a migrant tends to find many other migrants from the same rural area. The proximity of their origin villages naturally forms a close bound among these migrants; they tend to provide physical, psychological, or even financial support to one another. In a city far away from one's home, it is difficult for a migrant to find a similar supporting network. For this reason, a migrant would appear to give up some income in order to stay close to home.²² Empirically, it is rather challenging to isolate this social-network effect from the pure migration-distance effect, a topic better left for future research.

Our own interpretation is that this high income-distance elasticity is a result of the particular institutional context of rural-urban migration in China. Under the household responsibility system, all these migrant workers with rural *hukou* have access to some farm land in their home villages, which is allotted to each family based on headcount. When they

²¹Source of this information: <http://open.baidu.com/train/search>. It is the ticket price as of March 10, 2011; the price might be even lower in 2008 when our survey data were collected.

²²Of course this explanation assumes that earlier migrants tended to end up in nearby cities, which itself needs an explanation. Otherwise it is simply a circular argument.

migrate to ourban sectors, their family members, relatives, or neighbors may temporarily tend their farm. But the current land law in China does not allow the migrants to permanently transfer or sell their use rights to the land. In addition, there are other benefits tied to their rural *hukou*, which most of the migrant workers do not want to give up; and in order to keep such benefits, they may have to return immediately upon request from local authorities at their home villages. Due to the *hukou* system and other policy uncertainties, rural-urban migrants tend to consider their moves temporary and think of their home villages as a fallback place in case jobs in urban sector are not easily available (Yang, 1997; de la Rupelle et al., 2009). As a result, many of them have left their parents and children behind in home villages and thus they need to stay close in case any emergency occurs. All these suggest that an overhaul of the *hukou* system is the key to releasing rural-urban migrants from the strong hold of their home villages.

5 Conclusion

There has been a massive migration of population from rural to urban areas in China during the past three decades. We draw attention to the fact that rural-urban migrants in China prefer to live and work in cities close to their home villages, a tendency that helps explain some important rural-urban migration patterns in China. In this paper, we attempt to quantify the amount of income these migrants give up by staying close to home.

We build a simple model in which migrants from rural areas choose among potential destination cities to maximize utility. The distance between a destination city and the individual's home village is explicitly included in the utility function. Using some recent survey data, we first estimate an individual's expected income in each potential destination city by a semi-parametric method, controlling for potential self-selection biases. We then estimate the indirect utility function for rural-urban migrants in China. Our findings suggest that to induce an individual to migrate 10 percent further away from home, the wage paid to this migrant has to increase by 15 percent. This elasticity varies very little with distance; it is slightly higher for female than male migrants; it is not affected by the migrant's age, education, or marital status.

It remains unclear why rural-urban migrants in China so strongly prefer to stay close to home. We suspect that it has very much to do with the specific institutional context in China. Pinning down the exact reason behind this income-distance tradeoff is important. It not only helps us better understand the rural-urban migration patterns in China, but also has implications for urbanization policies in the country. We leave it for future work.

References

- [1] Bao, Shuming, Orn B. Bodvarsson, Jack W. Hou, and Yaohui Zhao (2009). “Migration in China from 1985 to 2000,” *Chinese Economy* 42, 7-28.
- [2] 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.
- [3] 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.
- [4] Chan, Kam Wing and Li Zhang (1999). “The Hukou System and Rural-Urban Migration in China: Processes and Changes,” *China Quarterly* 160, 818–855.
- [5] Dahl, Gordon (2002). “Mobility and the Return to Education: Testing a Roy Model with Multiple Markets,” *Econometrica* 70, 2367–2420.
- [6] Davies, Paul S., Michael J. Greenwood, and Haizheng Li (2001). “A Conditional Logit Approach to U.S. State-to-State Migration,” *Journal of Regional Science* 41, 337-360.
- [7] de la Rupelle, Maëlys, Quheng Deng, Shi Li, Thomas Vendryes (2009). “Land Rights Insecurity and Temporary Migration in Rural China,” IZA Discussion Paper No. 4668.
- [8] Falaris, Evangelos M. (1987). “A Nested Logit Migration Model with Selectivity,” *International Economic Review* 28, 429–443.
- [9] Lee, Lung-Fei (1983). “Generalized Econometric Models with Selectivity,” *Econometrica* 51, 507–512.
- [10] Lin, Justin Y., Gewei Wang, and Yaohui Zhao (2004). “Regional Inequality and Labor Transfers in China,” *Economic Development and Cultural Change* 52, 587-603.
- [11] McFadden, Daniel (1974). “Conditional Logit Analysis of Qualitative Choice Behavior,” in P. Zarembka, ed., *Frontiers in Econometrics*, Academic Press, New York, 105–142.
- [12] 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.
- [13] McFadden, Daniel and Kenneth Train (2000). “Mixed MNL Models for Discrete Response,” *Journal of Applied Econometrics* 15, 447–470.
- [14] Nakosteen, Robert A. and Michael Zimmer (1980). “Migration and Income: The Question of Self-Selection,” *Southern Economic Journal* 46, 840–851.

- [15] Poncet, Sara (2006). "Provincial Migration Dynamics in China: Borders, Costs and Economic Motivations," *Regional Science and Urban Economics* 36, 385–398.
- [16] Roback, Jennifer (1982). "Wages, Rents, and the Quality of Life," *Journal of Political Economy* 90, 1257-1278.
- [17] Robinson, Chris and Nigel Tomes (1982). "Self-Selection and Interprovincial Migration in Canada," *Canadian Journal of Economics* 15, 474–502.
- [18] Schwartz, Aba (1973). "Interpreting the Effect of Distance on Migration," *Journal of Political Economy* 81, 1153–1169.
- [19] Sinnott, R. W. (1984). "Virtues of the Haversine," *Sky and Telescope* 68, 159.
- [20] Timmins, Christopher (2009). "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.
- [21] Train, Kenneth (2009). *Discrete Choice Method with Simulation*, 2nd edition, Cambridge University Press, Cambridge, MA.
- [22] Yang, Dennis Tao (1997). "China's land arrangements and rural labor mobility," *China Economic Review* 8, 101–115.