

Zhang, Junfu; Zhao, Zhong

Working Paper

Measuring the income-distance tradeoff for rural-urban migrants in China

IZA Discussion Papers, No. 7160

Provided in Cooperation with:

IZA – Institute of Labor Economics

Suggested Citation: Zhang, Junfu; Zhao, Zhong (2013) : Measuring the income-distance tradeoff for rural-urban migrants in China, IZA Discussion Papers, No. 7160, Institute for the Study of Labor (IZA), Bonn

This Version is available at:

<https://hdl.handle.net/10419/69421>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

IZA DP No. 7160

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

Junfu Zhang
Zhong Zhao

January 2013

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

Junfu Zhang

*Clark University
and IZA*

Zhong Zhao

*Renmin University of China
and IZA*

Discussion Paper No. 7160
January 2013

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post Foundation. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ABSTRACT

Measuring the Income-Distance Tradeoff for Rural-Urban Migrants in China^{*}

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 migrant's home village and destination city 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 baseline estimates suggest that to induce a migrant to move 10 percent further away from home, the income of this migrant has to increase by 15 percent. This elasticity varies very little with migration 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.

JEL Classification: O15, R12, R23

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

Corresponding author:

Junfu Zhang
Department of Economics
Clark University
950 Main Street
Worcester, MA 01610
USA
E-mail: juzhang@clarku.edu

^{*} We thank Randy Akee, Gary Becker, Loren Brandt, Wayne Gray, James Heckman, Vernon Henderson, Albert Park, and Chih Ming Tan for their thoughtful comments. This paper has also benefited from comments and suggestions by seminar or conference participants at Clark University, BNU, UIBE, the 2nd CIER/IZA Workshop in Bonn, Germany, the 4th Migration and Development Conference at Harvard, the 6th IZA/World Bank Conference: Employment and Development in Mexico City, the Chicago-Renmin Symposium on Family and Labor Economics at University of Chicago, the Workshop on Labour Markets in China and North America at University of Western Ontario, and the 6th Meeting of the Urban Economics Association in Miami. Xiang Ao, Yue Chen, Kun Guo, Shuyi Lü, and Zehua Sun helped prepare the migration distance data. Alejandro Montero provided editorial assistance. Zhong Zhao would like to acknowledge financial support from the Peking University – Lincoln Institute Center for Urban Development. Collection of the Rural Urban Migration in China (RUMiC) data used in this paper is financed by IZA, ARC/AusAid, the Ford Foundation, and the Ministry of Labor and Social Security of China.

1 Introduction

China has a residence registration system, originally designed to control the movement of population 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 to be 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 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 140 million rural-urban migrants.¹

In this paper, we start with a brief overview of three stylized facts concerning the massive rural-urban migration in China. First, short-distance migration is more common than long-distance migration. Second, the earnings of rural-urban migrants vary substantially, depending on where they migrated; cities farther away from surplus labor generally pay higher wages to migrants. Third, labor-intensive industries initially in coastal regions are increasingly moving to inland China to take advantage of cheaper labor. We argue that the three stylized facts have to do with a simple phenomenon: In China, rural-urban migrants dislike moving far away from their home villages. The main goal of this paper is to quantify this preference for nearby migration destinations.

To provide a structural framework for our empirical analysis, we present a simple model in which migrants from rural areas choose among a set of destination cities to maximize utility. The distance between a migrant's home village and destination city 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 function for rural-urban migrants in China based on migration patterns. We try different specifications of the choice model, including the conditional logit, nested logit, and mixed logit. In all these specifications, we interact personal characteristics with migration distance and city characteristics to allow for heterogeneous preferences.

Our baseline estimates suggest that to induce a migrant to move 10 percent further

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.

away from home, the wage paid to this person has to increase by 15 percent. This elasticity varies very little with migration 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.

Researchers of internal migration have long recognized the tradeoff between income gain and migration distance (Sjaastad, 1962). Whereas we have not found any single study focusing exclusively on this issue, many authors have touched upon it in different contexts.² The common practice in the existing literature is to regress migration flows on distance, income level at destination, and other variables;³ the income-distance tradeoff is then inferred from the ratio of the coefficients of income and distance. Implicitly, this approach assumes that potential migrants expect to earn the same level of income as local residents at the destination, which is problematic as migrants may have observed and unobserved characteristics that are very different from those of local residents. Another problem with this traditional approach is that it generally ignores amenity differences between origin and destination locations. Given that people care about both earnings and local amenities, some may choose to migrate mainly to enjoy more amenities instead of to pursue higher earnings at the destination location. For this reason, without controlling for amenities, one would obtain biased estimates for both income and distance coefficients, resulting in an unreliable measure of the income-distance tradeoff.

In this paper, we attempt to improve upon this traditional approach with two key efforts, which constitute the main contribution of our study. First, methodologically, we adopt a new modeling framework. It not only deals more carefully with the estimation of migrants' income in potential destination locations, but also takes into account the amenity differences between destination and origin locations. Second, in terms of data and estimation, we make use of a unique survey on migrant households and therefore are able to estimate the income-distance tradeoff at the individual level. This allows us to relax assumptions researchers have to make when using aggregate data for model estimation; most importantly, we do not need to assume homogeneous preferences and are able to explore heterogeneity across individuals. Because of these improvements, we believe that we have obtained a more credible estimate of the income-distance tradeoff. This helps us better understand rural-urban migration patterns in China.

The rest of the paper is organized as follows. Section 2 discusses the three stylized facts on rural-urban migration in China. Section 3 presents a simple model of migration destination choice. Section 4 describes the data we use and the construction of key variables. Section 5 presents empirical results. Section 6 concludes.

²See, for example, Courchene, 1970; Davies et al., 2001; Fan, 2005; Phan and Coxhead, 2010; Poncet, 2006; Schwartz, 1973; and Vanderkamp, 1971.

³Normally, the income level at the origin is also included in the regression, either as a separate independent variable, or combined with the destination income to construct an independent variable (e.g., destination-origin income difference or ratio).

Table 1: Average monthly earnings of migrant household heads in fifteen major 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 migrating to a city), and city fixed effects. We then use the estimated coefficients to predict the average earnings in each city for a migrant with all independent variables set equal to the sample means.

2 Stylized Facts

Three stylized facts of China’s rural-urban migration have 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 the same or nearby provinces. Relatively fewer rural people in the far 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.⁴ Our own survey data on rural-urban migrants in 15 cities show that about half of them come from rural areas within the same province.

Second, the earnings of migrants vary substantially, depending on where they have migrated. Table 1 shows the average monthly earnings for rural-urban migrant household heads in 15 major destination cities. This average varies widely across cities. At one end of this distribution is Shanghai, where on average migrants make 2,338 yuan a month. At the

⁴Some 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 in China.

other extreme is Chongqing, where on average a migrant's income 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: Rural-urban migrants have very different income levels in different cities.⁵

Moreover, we note that in the Yangtze River Delta region, where the urbanization rate is relatively high and surplus rural labor is less abundant, migrants tend to earn more, as evidenced by the higher monthly income in cities such as Shanghai, Hangzhou, and Nanjing. In contrast, migrants tend to earn less in inland regions where surplus rural labor is plentiful, as in Chongqing, Luoyang, and Zhengzhou.

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 cheaper labor. 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.⁶ The most salient example is perhaps Foxconn, a contract manufacturer that makes products such as the iPod, iPad, and iPhone. It employs more than 400,000 migrant workers in the coastal city Shenzhen. In 2010, Foxconn announced a plan to construct new plants in inland cities such as Zhengzhou, Wuhan, and Chengdu, moving 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 home villages—helps explain these three stylized facts. As migrants tend to avoid long-distance moves, we observe shorter-distance migration more often. It is for the same reason that migrant earnings are far from being equalized across cities; cities with limited surplus labor in nearby rural areas have to offer higher wages to attract migrant workers from remote regions. Similarly, this reluctance to move far away requires employers in distant regions to pay higher wages, which motivates labor-intensive industries to move toward surplus labor. Originally, labor intensive industries, especially 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 China has improved substantially.⁷ As a result, the cost of recruiting migrant workers has become a more prominent factor in firms' location

⁵We are by no means suggesting that migrant income will be equalized across cities if people do not care about migration distance. As is well known, even under zero moving costs, identical individuals in different cities may earn different levels of income in equilibrium, simply because amenities and cost of living vary across cities (Roback, 1982). But we suspect that the variation of migrant earnings in China cannot be fully explained by such differences.

⁶See <http://finance.ifeng.com/roll/20100917/2631649.shtml> (viewed on February 19, 2011).

⁷The greatly improved transportation infrastructure in China and its impact on the distribution of population and economic activities have recently stimulated considerable research interest. See, for example, Banerjee et al. (2012), Baum-Snow et al. (2012), and Baum-Snow and Turner (2012).

decisions, which explains why the “inward moving wave” started only recently.

There are many possible reasons as to why rural-urban migrants prefer closer migration destinations. When an individual migrates to a city far from her village, she will be disconnected from her social-family network, a most reliable source of emotional, physical, psychological, and even occasional financial support in rural communities. She may have to live in an unfamiliar environment with different weather, food, and culture. She may feel isolated, insecure, and worry about discrimination. For all of these reasons, one would be willing to give up some income in order to stay closer to their home village. 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.

In the next section, we present a simple model of migration destination choice, which provides a framework for empirical analysis and interpretation of results.

3 A Model of Migration Destination Choice

3.1 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.⁸ If living in city j , individual i faces the following utility-maximization problem

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

- 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 .⁹

- 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 .

⁸Following Bayer et al. (2009) and Timmins (2007), in our empirical analysis we focus on household heads only, assuming that they are the decision makers.

⁹We use “non-tradable” to describe any goods or services that have different prices in different cities. In addition to housing, many other goods are non-tradable across cities in China; this 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 for 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 migration distance to cause 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 $\delta \equiv \left(\frac{\alpha_C}{\alpha_C + \alpha_H} \right)^{\alpha_C} \left(\frac{\alpha_H}{\alpha_C + \alpha_H} \right)^{\alpha_H}$ and $\alpha \equiv \alpha_C + \alpha_H$. Rescaling by $\frac{1}{\delta}$, we rewrite the indirect utility function as

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

Our modeling framework suggests two natural ways to measure the income-distance tradeoff. First, we could look at the amount of money individual i is willing to give up in order to live closer to her home village by one unit of distance, i.e., person i 's willingness to pay for a closer destination. From equation (2), we calculate the marginal rate of substitution between migration distance and income:

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

which can be interpreted as the marginal willingness to pay. It is higher when i has a higher income, and lower when i is further way from home. Moreover, this willingness to pay is higher when the the ratio $\frac{\beta}{\alpha}$ is higher.

Second, we could also examine the percentage increase in income needed to compensate an individual in order for her to migrate one percent further away from home, i.e., the income-distance elasticity. Taking the natural log of equation (2) and holding the utility level constant, we see that this income-distance elasticity is exactly $\frac{\beta}{\alpha}$:

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

That is, to induce an individual to migrate one percent further away from home, it is necessary to offer this person an income that is $\frac{\beta}{\alpha}$ percent higher.

Either way, α and β are the two key parameters needed to measure the income-distance

tradeoff. Thus our main task in this paper is to empirically estimate α and β , so that we can calculate a migrant's willingness to pay for a closer destination and the income-distance elasticity.

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:

$$\ln I_{ij} = Z_i \hat{\gamma}_j + \varepsilon_{ij}. \quad (3)$$

Z_i is individual i 's characteristics that affect expected earnings, including for example age, gender, education, and marital status. $\hat{\gamma}_j$ is a set of city-specific coefficients that determine how individual characteristics are rewarded in city j . We need to control for potential self-selection biases when estimating $\hat{\gamma}_j$; 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.¹⁰ 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

$$\ln V_{ij} = \alpha(Z_i \hat{\gamma}_j) - \beta \ln D_{ij} + \theta_j + v_{ij}, \quad (5)$$

where $\theta_j \equiv g(X_j) - \alpha_H h(X_j) - \alpha_H \epsilon_j + \xi_j$ and $v_{ij} \equiv \alpha \varepsilon_{ij} + \eta_{ij}$. 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

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

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

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

where κ_{ij} is an indicator function that equals 1 if individual i is observed in city j . We can estimate $\{\alpha, \beta, \theta_1, \dots, \theta_J\}$ by maximizing this likelihood function.¹¹ Note that if any

¹⁰One can easily derive a relationship like this in a Rosen-Roback type model. See, e.g., Roback (1982).

¹¹The conditional logit approach is commonly used for the analysis of migration choice. See, for example,

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 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 .¹²

To avoid cluttering notations, we have thus far treated β as a constant; we have also dumped the effects of both observed and unobserved city characteristics into the city fixed effect, forcing everybody's utility from the characteristics of city j to be the same θ_j . In empirical specifications below, we shall relax these assumptions and use more flexible function forms. We will allow β to vary with distance or individual characteristics. We will also allow the preference for observed city characteristics X_j to vary across individuals and take the differential effects out of the city fixed effect.

3.2 Empirical specifications of the model

3.2.1 Nonconstant disutility of migration distance

The distaste for migration distance (β) is not necessarily a constant. In our empirical specifications, 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 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:

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

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.

¹²Conceptually, 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_j) as well as the cost of maintaining such characteristics (supply of X_j).

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 into 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.

3.2.2 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 heterogeneity may bias our estimates of the key parameters β and α . Thus we also experiment with alternative specifications that account for differential preferences.

Specifically, we assume that individual i 's utility from city j 's K different characteristics, $X_j \equiv (X_j^1, \dots, X_j^K)'$, is given by

$$\begin{aligned}\Omega_{ij} &= \bar{c}_j + \sum_{k=1}^K \left[c_{1k} (G_i X_j^k) + c_{2k} (A_i X_j^k) + c_{3k} (E_i X_j^k) + c_{4k} (M_i X_j^k) \right] \\ &= \bar{c}_j + S_i C X_j,\end{aligned}\tag{9}$$

where $S_i \equiv (G_i, A_i, E_i, M_i)$ are the same individual characteristics as defined above, C is a $4 \times K$ matrix of coefficients, and \bar{c}_j is the average utility derived from all these characteristics of city j . We can therefore rewrite equation (5) as

$$\ln V_{ij} = \alpha (Z_i \hat{\gamma}_j) - \beta \ln D_{ij} + S_i C X_j + \bar{\theta}_j + v_{ij},\tag{10}$$

where we have replaced θ_j with $S_i C X_j + \bar{\theta}_j$. This $\bar{\theta}_j$ is still a city fixed effect, which captures \bar{c}_j as well as utilities from unobserved city characteristics. We can now estimate the parameters by maximizing the following likelihood function

$$\tilde{L} = \prod_i \prod_{j=1}^J \left\{ \frac{\exp[\alpha(Z_i \hat{\gamma}_j) - \tilde{\beta} \ln D_{ij} + S_i C X_j + \bar{\theta}_j]}{\sum_{s=1}^J \exp[\alpha(Z_i \hat{\gamma}_s) - \tilde{\beta} \ln D_{is} + S_i C X_s + \bar{\theta}_s]} \right\}^{\kappa_{ij}},$$

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

It's worth noting that allowing for differential preferences for city characteristics in equation (10) plays a key role in identification. Since we have to use predicted instead of actual income to estimate parameter α in the utility function and individual characteristics Z_i appear in the predicted income ($Z_i \hat{\gamma}_j$), model identification requires that some individual

characteristics in Z_i do not enter the utility function as separate linear terms. Our basic setup in equation (5) obviously meets this requirement, since it has excluded all individual characteristics from the utility function. However, such a specification is rather unrealistic. Suppose that educational attainment is one of the variables in Z_i used to predict income. The utility function in equation (5) assumes that more educated migrants do not derive any more (or less) utility from any destination city, which is clearly a strong assumption. By adding the interaction terms $S_i C X_j$ in the utility function, we have relaxed this assumption in equation (10) and still achieve identification. Now more educated migrants can derive more (or less) utility from a destination city; identification is instead based on the assumption that more educated migrants derive more (or less) utility from a destination city only because the city has some characteristics that are more (or less) attractive to educated people. More generally, the identification of parameters in equation (10) only imposes the following restriction: If certain types of individuals value a city more than others, it is only because such individuals care more about the observed characteristics of the city included in X_j . Given that we will interact four individual characteristics with nine different city characteristics in our baseline regressions and that we will add even more interaction terms in our sensitivity analysis, we believe that this identification restriction is plausible.

3.2.3 Nested logit

Although the conditional logit model is the standard approach to estimating migration choices, we would like to check whether our findings are sensitive to this specification. It is well known that the conditional logit model assumes the *independence from irrelevant alternatives* (IIA).¹³ This assumption might not hold given that some destination cities in our data are physically close to one another. For example, Dongguan, Shenzhen, and Guangzhou all belong to Guangdong province and are all in the Pearl River Delta region. These cities share some common unobserved characteristics such as similar weather and the same dialect, which will likely cause violation of IIA. As a solution, we try an alternative specification—the standard nested logit model.

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

$$\begin{aligned}
 \ln V_{ij} &= \alpha (Z_i \hat{\gamma}_j) - \beta \ln D_{ij} + \sum_{k=1}^K \left[c_{1k} (G_i X_j^k) + c_{2k} (A_i X_j^k) + c_{3k} (E_i X_j^k) + c_{4k} (M_i X_j^k) \right] \\
 &\quad + \sum_{s=1}^J \bar{\theta}_s \kappa_{is} + v_{ij} \\
 &= \Psi_{ij} \Upsilon + v_{ij},
 \end{aligned} \tag{11}$$

¹³Let 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 .

where

$$\Psi_{ij} \equiv ((Z_i \hat{\gamma}_j), -\ln D_{ij}, G_i X_j^1, A_i X_j^1, E_i X_j^1, M_i X_j^1, \dots, G_i X_j^K, A_i X_j^K, E_i X_j^K, M_i X_j^K, \kappa_{i1}, \dots, \kappa_{iJ})$$

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

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 .¹⁴ 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, parameters in 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} is equal to one if i chooses city j and j is in region n , and zero otherwise.

3.2.4 Mixed logit

Although we allow β to vary, we have imposed stringent functional-form restrictions on how it varies. To check the sensitivity to these restrictions, we will estimate a mixed logit model. In this alternative specification, we treat the two key parameters, β and α , as random variables across individuals.¹⁵ We assume that they follow a specific joint distribution but impose nothing on how each parameter varies across individuals. Once the distribution of β and α are estimated, we use their mean values to calculate the income-distance elasticity.

We again specify the indirect utility function as in equation (11), allowing for heterogeneous preferences for observed city characteristics:

$$\ln V_{ij} = \Psi_{ij} \tilde{\Upsilon} + v_{ij}. \quad (12)$$

¹⁴As is well known, this nested logit model reduces to the standard logit model if $\lambda_n = 1 \forall n$ (McFadden, 1978).

¹⁵The 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 all city characteristics are unique to each city, one has to drop some city dummies in order to add those city characteristics; otherwise, there will be perfect collinearity.

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
- (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 $\tilde{\Upsilon}$.¹⁶

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. We 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 the 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 the income-distance elasticity.

¹⁶We 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 12) 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 identical to 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).

4 Data and Key Variables

For our 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. With funding from various sources, these researchers are conducting a five-year longitudinal survey in China and Indonesia, with the goal of studying 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, collected in 2008. 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, we use data 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.

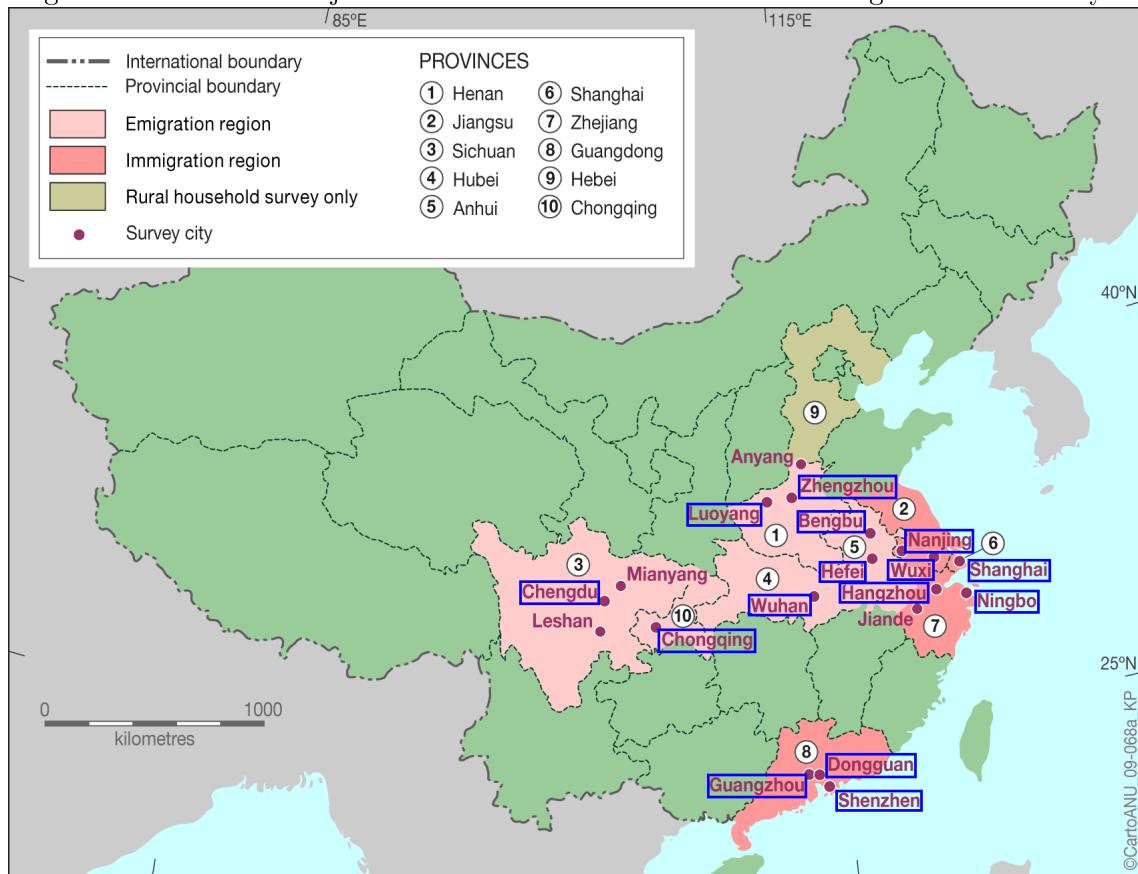
The migrants surveyed were randomly chosen from 15 cities that are major urban destinations for rural migrants in China.¹⁷ Eight of these cities are in coastal regions (Shanghai, Nanjing, Wuxi, Hangzhou, Ningbo, Guangzhou, Shenzhen, and Dongguan); five are in central inland regions (Zhengzhou, Luoyang, Hefei, Bengbu, and Wuhan); and two 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 estimate the income-distance tradeoff with high precision.

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 about a household member, we know the person's age, gender, education level, current address, home address before migration, etc. Regarding employment experience, we know the person's occupation, monthly income, how he/she found the current job, whether the person is self-employed or a wage worker, what was his/her first job, how he/she found the first job, etc. Some of these variables will be

¹⁷The RUMiC survey group first identified the top four destination provinces and the top five origin provinces for rural-urban migrants, and then selected 15 major cities in these provinces as their survey sites. A 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 Kong (2010) and the RUMiC survey group's homepage (<http://rse.anu.edu.au/rumici/documentation.php>) for detailed documentation of the sampling method.

Figure 1: The fifteen major destination cities where rural-urban migrants were surveyed



Source: The Rural-Urban Migration in China and Indonesia Project Website (<http://rse.anu.edu.au/rumici/pdf/china.pdf>), with modifications by the authors. 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.

useful when we explore various factors that help explain our main findings.

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 ($\ln \hat{I}_{ij} = Z_i \hat{\gamma}_j$), 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-line data source, we find the latitude-longitude coordinate for each home county and each destination city.¹⁸ We then use the Haversine formula to calculate the “great-circle distance” (on the surface of the Earth) from the home county to each city.¹⁹

In theory, physical, cultural, and social distances perhaps all matter in a person’s migration decision. Here we use the physical distance only and assume that other relevant distances are highly correlated with physical distance.²⁰ Even for physical distance, one might argue that railway or highway distance is more appropriate. However, such information at the county level is difficult to obtain and changes almost daily because China has been continuously upgrading its transportation infrastructure (Baum-Snow et al., 2012; Baum-Snow and Turner, 2012). We therefore use the “great-circle distance” as a proxy.

To generate expected income $\ln \hat{I}_{ij}$, we run a series of city-specific regressions of income on individual characteristics, from which we derive estimated coefficients 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).²¹

¹⁸The online data source is <http://ngcc.sbsm.gov.cn/Mapquery/default.aspx>, the website of the National Geomatics Center of China.

¹⁹Let $(lat_j, long_j)$ 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.

²⁰We have looked into possible ways to measure cultural distance. There are indeed such measures for China constructed by geographers, but they are based on physical distance. In other words, they are just adjusted physical-distance measures. Given that cultural distance measures are generally discrete and hard to interpret, we decided to stick to the physical distance.

²¹It 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

Consider the following model

$$\ln I_{ij} = Z_i \gamma_j + \varepsilon_{ij},$$

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

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 ; $\psi(\cdot)$ is an unknown function that gives the conditional mean $E(\varepsilon_{ik}|\cdot)$; and e_{ij} represents the remaining error. Dahl (2002) introduces an “index sufficiency assumption” which assumes 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 of where she actually migrated.

The key to implementing Dahl’s method is to nonparametrically estimate the probability of each individual migrating to her destination 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.²² 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.²³ For each individual i in city j , we find the cell she belongs to. The

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).

²²It 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. Therefore the group is too small to be treated as a separate one.

²³There is a tradeoff between the number of cells and the precision of the estimated migration probability. Because each individual can choose among 15 different destination cities, we need a reasonably large number

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} :

$$\ln I_{ij} = Z_i \gamma_j + \psi_{j1} \hat{P}_{ij} + \psi_{j2} (\hat{P}_{ij})^2 + \mu_{ij}. \quad (13)$$

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

A few remarks are in order about the specification in equation (13). First, a polynomial function also has a constant, but we cannot include it in the regression because of the constant in Z_i . It is impossible to separately identify both of them. However, note that it is not crucial to accurately estimate the constant term in this income regression. Suppose everybody's predicted income in city j is biased upward by a constant amount. Because we have included a city fixed effect in the logit regression, this will only change the city fixed effect and will not bias the estimation of α .

Second, we have included self-employment status in this regression, which requires some explanation. In our sample, about twenty percent of the migrants are self-employed in urban areas and their average earnings are substantially higher than wage workers. It is clear that these migrants have some unobserved characteristics; they may have high abilities, more financial capital, or a different attitude toward risks, which enable them to earn more through self-employment. Our specification here essentially uses the self-employment status as an indicator to capture such unobserved qualities. By doing so, we assume that the currently self-employed will only look at the self-employed in a city to form their expectations of earnings in the city and similarly wage workers only look at wage workers. We need this simplifying assumption because dealing with self-selection along two dimensions (across different cities as well as different employment status) is much more data demanding. We experimented with a specification of the income equation that excluded self-employment status, estimated using the whole sample or only wage workers. In either case, the predicted income has a much smaller coefficient in the logit regression and the coefficient is estimated with much less precision, suggesting that the predicted income contains too much noise. Since economic theory predicts that expected income matters in migration decisions, we used theory as our guide and decided to work with this specification as it appears to predict income more accurately.²⁴

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.

²⁴The theoretical claim that expected income drives migration is confirmed by a wide range of existing

Third, we have made an implicit assumption that given the personal characteristics included in Z_i , a migrant's home province does not affect her or his earnings in city j . In other words, two migrants with identical personal characteristics working in the same city, but coming from different provinces, are expected to earn the same amount. Since we have used education levels and home province to predict migration probabilities and education naturally enters the income equation, identification requires that home province can be excluded from equation (13). The migrants in our sample mostly do manual work in factories, construction sites, or low-skill service industries. Their cultural backgrounds, language abilities, and inter-personal skills, which are likely to vary across home provinces, are generally unimportant for their job performance. Therefore, it is perhaps not a strong assumption that earnings do not directly depend on home province. As an informal test, for each city we calculate the residuals from the income regression and regress it on the whole set of 29 home province dummies, to check whether home province dummies can explain a large portion of the unaccounted variations in income. We find that all these regressions have very small R^2 , never higher than 0.085. For each regression, we test the hypothesis that the coefficients of home province dummies are jointly equal to zero. Even at the 20-percent significance level, we cannot reject this hypothesis in all but one case (for the city of Guangzhou). Indeed, for 12 out of the 15 cities, the p-value of this F test is higher than 0.5. Overall, these exercises suggest that it is reasonable to exclude a migrant's home province from the earnings regressions for these cities.

Besides estimated income, we have also collected information on destination city characteristics from the *Urban Statistical Yearbook of China*.²⁵ 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. In our logit regressions, we interact these variables with individual characteristics to allow for differential preferences for observed city characteristics.

5 Empirical Results

We present empirical results in this section.

5.1 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 descriptive statistics in Table 2.

studies. See for example Kennan and Walker (2011) for a recent study that focuses on U.S. internal migration.

²⁵We 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 (yuan)	1,758.67	2,508.09	0	99,998
Self-employed	0.225	0.418	0	1
Log migration distance	5.364	1.153	1.557	8.309
From same 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.

Seventy-one percent of these migrants are male; 61 percent are married. The average migrant 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 migrants are intra-province migrants.

Among these migrants, 22.5 percent are self-employed. These individuals can be restaurant owners, convenience store owners, scrap metal collectors, street vendors who sell fruits, snacks, cigarettes, clothing, souvenirs, etc. or provide services such as shining shoes and repairing bicycles or electronics. A large proportion of the self-employed migrants work alone; only a quarter of them (25.4 percent) also hire other people. Among those who hire other people, the average number of employees is 3.5. Self-employed migrants on average earn more income than wage workers.²⁶

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 the entire period. There might be some time in between when the migrant returned to the home village and then migrated out again. 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 initial destination provinces. That is, a migrant might first migrate to province A, but later found a better job opportunity in province B and then 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 arise, which is important because we model them as utility maximizers.

²⁶See Zhang and Zhao (2012) for a study that focuses on these self-employed migrants.

5.2 Regression results

We start by presenting in Table 3 the results from the earnings regressions for the 15 destination cities. Age always has a positive coefficient, which is statistically significant in most cases. That is, older migrants generally earn more. This effect is nonlinear and falls off eventually, as implied by the negative coefficient of age squared. Male migrants always earn more; more educated migrants almost always earn more; and self-employed migrants always earn more. Marital status, on the other hand, is not significantly correlated with earnings except in one city. Whereas the qualitative results are quite similar across the 15 destination cities, the magnitude of the coefficients varies a great deal. For example, the male dummy has a coefficient of 0.546 in Hefei, in contrast to 0.015 in Dongguan. Similarly, while an extra year of schooling increases log monthly earnings by 0.046 in Guangzhou, its effect is only half as large in Chongqing.

The estimated migration probability has a (marginally) significant coefficient in the regressions for Guangzhou and Dongguan; for all other cities, neither the migration probability nor its squared term has a significant coefficient. We conduct a Wald test with a null hypothesis that these two coefficients are jointly equal to zero, and cannot reject this hypothesis except in the city of Dongguan. Therefore, despite our concerns over self-selection in income regressions, it does not seem to be a serious problem in our sample.

We next show results from the conditional logit regressions in Table 4. The specification in column (1) includes log income, log distance, city fixed effects, and allows for differential preferences for observed city characteristics. We have four individual characteristics (gender, age, education, and marital status) and nine city characteristics (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 regressions.²⁷ The coefficient is 1.391 for log income and 2.101 for (negative) log distance, giving an estimated $\frac{\beta}{\alpha}$ as 1.511. This estimate implies that income has to increase by 15 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 except Shanghai and Ningbo, suggesting that, if not for income and distance reasons, most other

²⁷Note that we do not interact self-employment status with city characteristics here, although it is included in the earnings regressions. That is, we are assuming the following: Once we take into account the effects of age, gender, education, marital status, and higher expected earnings, being self-employed does not make a person value a city more or less than wage workers. This seems plausible to us. We tried to relax this assumption by adding the interaction terms between self-employment status and city characteristics. This led to a smaller and less precisely-estimated α but had almost no effect on the estimated β in the logit regressions. Given that a smaller α gives a higher estimate of $\frac{\beta}{\alpha}$ and biases against our key findings, we decided to take the conservative approach and not interact self-employment status with city characteristics.

Table 3: OLS regression results on log monthly earnings in destination cities

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
	Guangzhou	Dongguan	Shenzhen	Zhengzhou	Luoyang	Hefei	Bengbu	Chongqing	Shanghai	Nanjing	Wuxi	Hangzhou	Ningbo	Wuhan	Chengdu
Constant	5.905 (0.228)	6.506 (0.279)	5.762 (0.295)	5.167 (0.389)	6.724 (0.768)	6.297 (0.806)	5.126 (0.432)	5.695 (0.320)	6.167 (0.446)	6.510 (0.274)	7.274 (0.355)	6.353 (0.317)	5.313 (0.596)	6.022 (0.524)	5.726 (0.360)
Age	0.056 (0.013)	0.037 (0.015)	0.083 (0.017)	0.081 (0.019)	0.013 (0.041)	0.014 (0.040)	0.092 (0.023)	0.057 (0.018)	0.041 (0.025)	0.038 (0.015)	0.003 (0.020)	0.041 (0.019)	0.082 (0.033)	0.041 (0.030)	0.044 (0.020)
Age squared	-0.0008 (0.0002)	-0.0006 (0.0002)	-0.0013 (0.0002)	-0.0011 (0.0002)	-0.0003 (0.0005)	-0.0003 (0.0005)	-0.0013 (0.0003)	-0.0009 (0.0003)	-0.0006 (0.0003)	-0.006 (0.0002)	-0.0000 (0.0003)	-0.0007 (0.0003)	-0.0011 (0.0004)	-0.0006 (0.0004)	-0.0008 (0.0003)
Male	0.190 (0.037)	0.015 (0.050)	0.065 (0.046)	0.309 (0.072)	0.352 (0.132)	0.546 (0.163)	0.053 (0.114)	0.320 (0.053)	0.219 (0.066)	0.147 (0.049)	0.147 (0.051)	0.182 (0.061)	0.265 (0.130)	0.300 (0.086)	0.259 (0.063)
Years of schooling	0.046 (0.008)	0.036 (0.010)	0.028 (0.010)	0.042 (0.020)	-0.012 (0.044)	0.031 (0.035)	0.049 (0.020)	0.023 (0.013)	0.042 (0.014)	0.016 (0.012)	-0.011 (0.018)	0.026 (0.011)	0.035 (0.020)	0.028 (0.030)	0.030 (0.014)
Married	0.039 (0.051)	-0.023 (0.057)	0.059 (0.065)	0.020 (0.106)	-0.050 (0.211)	0.113 (0.193)	0.032 (0.116)	0.034 (0.074)	-0.036 (0.090)	0.051 (0.078)	0.077 (0.079)	0.101 (0.078)	-0.177 (0.181)	-0.070 (0.122)	0.223 (0.097)
Self-employed	0.388 (0.072)	0.190 (0.092)	0.282 (0.101)	0.162 (0.078)	0.744 (0.142)	0.653 (0.129)	0.492 (0.087)	0.308 (0.073)	0.581 (0.076)	0.116 (0.068)	0.319 (0.075)	0.305 (0.071)	-0.188 (0.150)	0.524 (0.102)	0.650 (0.075)
$\hat{P}_{i,j}$	-0.904 (0.519)	-4.006 (2.160)	0.148 (1.580)	0.896 (1.032)	7.207 (9.405)	-1.386 (10.91)	-3.743 (3.424)	0.513 (1.104)	-0.962 (2.005)	0.173 (0.608)	0.261 (1.370)	0.428 (0.657)	4.369 (4.549)	-0.440 (1.519)	2.004 (3.187)
$(\hat{P}_{i,j})^2$	1.090 (0.660)	13.27 (9.358)	1.351 (6.550)	-2.444 (2.008)	-34.58 (35.80)	1.732 (31.02)	16.26 (16.09)	-0.577 (1.136)	2.589 (6.882)	-0.170 (1.100)	0.265 (3.736)	-0.442 (0.956)	-14.13 (19.45)	0.624 (2.088)	-2.965 (5.057)
p-value of test $(\hat{P}_{i,j} = 0)$ and $((\hat{P}_{i,j})^2 = 0)$.	0.183	0.011	0.413	0.267	0.107	0.307	0.501	0.730	0.838	0.824	0.165	0.521	0.314	0.954	0.638
R^2	0.251	0.134	0.172	0.168	0.211	0.141	0.381	0.188	0.151	0.088	0.144	0.132	0.089	0.118	0.326
Number of observations	394	297	297	342	198	338	193	390	488	377	200	389	196	395	386

Standard errors are in parentheses. To check for the existence of selection biases, we test the null hypothesis that the coefficients of $\hat{P}_{i,j}$ and $(\hat{P}_{i,j})^2$ are jointly equal to zero.

cities are less attractive to migrants than Guangzhou. The difference is the largest for Luoyang, Chengdu, and Zhengzhou, all inland cities in less developed regions. All of these results seem to make sense. We examine the simple correlation between the city fixed effects and observed city characteristics. We find that these 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 4, we estimate different values of β for different quartiles of migration distance. They are more or less the same, ranging from 2.068 to 2.117. Because these parameters are so precisely estimated, it turns out that 2.068 is statistically significantly different from 2.117. 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 quartiles. They are all close to 1.5. Therefore, it appears that income-distance elasticity changes very little with distance, which is somewhat surprising.

In column (3) of Table 4, we allow β to vary with individual characteristics by adding the interaction terms between log distance and individual characteristics. This regression shows that 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 1.571 for female migrants, in contrast to 1.454 for male migrants. In other words, it is a bit easier to induce male than female migrants to move further away from home.

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 care less about elementary schools, perhaps because they no longer have school-aged children.

We also tried three specifications parallel to those in Table 4, where the only difference is that all the interaction terms between individual and city characteristics were excluded from the regressions (not presented here). Our goal was to explore whether it makes a difference when allowing for differential preferences over all observed city characteristics. It turns out that the extra control variables only affect $\hat{\alpha}$, the estimated coefficient of log income. This estimate is always lower when not controlling for observed city characteristics: all below 1.1 as opposed to the estimates close to 1.4 in Table 4. Excluding city level controls still gives a coefficient of log distance close to 2, and therefore the estimated $\frac{\beta}{\alpha}$ is higher, at about 2.0. This suggests that city level controls may affect our result substantially, and therefore we will explore specifications with more controls in sensitivity analysis.

Table 4: Conditional logit regression results

Variable coefficient name	(1) Conditional Logit	(2) Conditional Logit	(3) Conditional Logit
Log income α	1.391 (0.380) [0.263]	1.396 (0.374) [0.264]	1.414 (0.385) [0.267]
-Log distance β or b_0	2.101 (0.088) [0.040]		2.123 (0.305) [0.254]
-Log distance*1 $_{Q1}$ β_1		2.087 (0.099) [0.052]	
-Log distance*1 $_{Q2}$ β_2		2.117 (0.104) [0.047]	
-Log distance*1 $_{Q3}$ β_3		2.103 (0.088) [0.045]	
-Log distance*1 $_{Q4}$ β_4		2.068 (0.084) [0.045]	
-Log distance*male b_1			-0.165 (0.081) [0.093]
-Log distance*age b_2			0.002 (0.005) [0.006]
-Log distance*education b_3			0.002 (0.024) [0.018]
-Log distance*married b_4			0.020 (0.085) [0.106]
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 (0.449) [0.849]	1.494 (0.431) [0.845] 1.516 (0.442) [0.858] 1.506 (0.438) [0.855] 1.481 (0.424) [0.840]	Female: 1.571 (0.427) [0.846] Male: 1.454 (0.396) [0.791]

Robust standard errors clustered by city are in parentheses; bootstrapped standard errors are in brackets. 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, and per capita air pollutants), and therefore a total of 36 interactions. There are 4,434 migrant household heads between 20 and 60 years old, 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.

Before moving on, we remark on two issues. The first concerns the standard errors of our estimates in Table 4. A key feature of our empirical approach is that we use estimated regressors. In the income regressions, we use estimated migration probabilities to correct for potential selection biases; in the logit regressions, we use estimated earnings instead of actual earnings. It is well known that estimated regressors could bias standard errors of the estimates computed in the normal way. For simple two step models, there exist analytical solutions to this problem (Murphy and Topel, 1985). However, such solutions are not applicable here because our approach is more complicated. We have three steps; we have used both parametrically and nonparametrically estimated regressors; and our second step involves 15 parallel income regressions.²⁸

Following common practice, we bootstrap standard errors for our conditional logit regressions to estimate the true standard errors and assess the biases in the normally computed standard errors.²⁹ Specifically, for each destination city, we first draw a random sample from our data (with replacement) of the same size as the real sample. We then perform our three-step estimation using these bootstrapped data: (1) estimate migration probability P_{ij} nonparametrically; (2) estimate 15 income equations, correcting for potential selection biases, and predict $\ln \hat{I}_{ij}$ for all i and j ; and (3) estimate the logit model and calculate $\frac{\beta}{\alpha}$. This process is repeated 1000 times. For each coefficient in the logit model, the standard deviation of the 1000 estimates is taken as the bootstrapped standard error.

We also compute non-robust standard errors and robust standard errors clustered by city in the normal way. For each key parameter, the non-robust standard error is always the smallest among the three; in most cases, the clustered standard error is the largest and the bootstrapped standard error is somewhere in between.³⁰ We have reported the two larger (bootstrapped and clustered) standard errors in Table 4 for comparison purposes. Using either standard errors, the key parameters are always statistically significant. Because the clustered standard errors are usually larger than the bootstrapped standard errors in these baseline regressions and because bootstrapping is computationally intensive, in our sensitivity analysis below, we will only report the clustered standard errors.

The second remark is about our practice of estimating model parameters using a migrant sample from 15 major destination cities only. In reality, any migrant in our sample almost surely has more than 15 possible destinations, making one wonder whether our estimates are biased. It is important to note here that this creates no problem in our baseline estimations. McFadden (1978) has shown that maximization of the conditional-logit likelihood function

²⁸The asymptotic variance of estimators in multiple-step regressions with generated regressors is a topic of on-going research. See, e.g., Hahn and Ridder (2010).

²⁹In a special case, which involves estimating mixed logit regressions with generated regressors, Petrin and Train (2010) are able to compute both analytical standard errors and bootstrapped standard errors. They find them quite similar.

³⁰Whereas the bootstrapped standard errors are smaller than clustered standard errors for the estimates of α and β , they are larger for the estimates of $\frac{\beta}{\alpha}$. This is because the estimates of α and β from bootstrapped samples show a correlation close to zero, but those estimated using the real sample are positively correlated (and thus their ratio has a smaller variance).

based on a subset of alternatives still produces consistent estimates, as long as the so-called “uniform conditioning property” holds. Let $P(J|k)$ be the probability that the researcher will include the subset of alternatives J in the regression conditional on alternative k being chosen by the decision maker. The uniform conditioning property states that $P(J|k)$ is the same for all k in J . Given that our 15 destination cities were chosen by the RUMiC survey team based on overall migration patterns and thus independently of each individual migrant’s choice, the uniform conditioning property holds trivially. Therefore, at least in our baseline conditional logit regressions, the results can be interpreted as consistent estimates for the migrants in our sample. Further, since the sample in each of the 15 cities is representative, our baseline estimates imply that migrants in these cities have an income-distance elasticity close to 1.5.³¹

However, for nested and mixed logit models, the uniform conditioning property does not guarantee consistency, therefore estimates based on a subset of alternatives are not easily interpretable. Given this problem, we consider the conditional logit models our preferred specification and estimate the nested and mixed logit models mainly for sensitivity analysis.

In Table 5, we present results from nested logit regressions. In China, the Pearl River Delta and the Yangtze River Delta are the two leading commercial and manufacturing regions. 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; and 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, as in Table 4.

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 unlikely to hold. However, the alternative nested logit specification has very limited effect on our key estimates. The estimated $\frac{\beta}{\alpha}$ is slightly larger, but still close to 1.5. It does not vary much across different distance quartiles. The gender of the migrants still makes a difference: Whereas the ratio is 1.638 for females, it is 1.497 for male migrants. Therefore, although it seems desirable to

³¹There remains a crucial question whether these migrants observed in the 15 major destination cities represent the whole population of potential and actual rural-urban migrants in China. It is perhaps reasonable to speculate that migrants in smaller cities, return migrants, and individuals who have never moved out of rural areas have even higher income-distance elasticities, yet we cannot confirm this due to data constraints. While the RUMiC research group also surveyed some rural communities, they focused on targeted areas and collected information on local rural population and return migrants only, which cannot be used to perform similar estimations. A more representative study would have to rely on population census data. However, the most recent census in China did not ask about income at all, which is not very useful for estimating the income-distance tradeoff.

Table 5: Nested logit regression results

Variable coefficient name	(1) Nested Logit	(2) Nested Logit	(3) Nested Logit
Log income α	1.267 (0.419)	1.316 (0.429)	1.291 (0.429)
-Log distance β or b_0	1.989 (0.098)		1.964 (0.299)
-Log distance* 1_{Q1} β_1		2.007 (0.141)	
-Log distance* 1_{Q2} β_2		2.043 (0.139)	
-Log distance* 1_{Q3} β_3		2.020 (0.122)	
-Log distance* 1_{Q4} β_4		1.995 (0.121)	
-Log distance*male b_1			-0.181 (0.084)
-Log distance*age b_2			0.005 (0.005)
-Log distance*education b_3			0.003 (0.024)
-Log distance*married b_4			-0.046 (0.089)
Differential preferences for city characteristics	Yes	Yes	Yes
City fixed effects	Yes	Yes	Yes
Log likelihood	-6,550.66	-6,541.60	-6,547.26
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 $\frac{\beta}{\alpha}$	1.570 (0.526)	1.525 (0.495) 1.553 (0.509) 1.535 (0.503) 1.516 (0.491)	Female: 1.638 (0.545) Male: 1.497 (0.498)

Robust standard errors clustered by city 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 a more detailed explanation). There are 4,434 migrant household heads between 20 and 60 years old, 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 6: Mixed logit regression results

Variable coefficient name	(1) Mixed Logit Jointly normal α, β	(2) Mixed Logit Jointly log-normal α, β
Log income α	Mean: 1.702 (0.420) Standard deviation: 1.384 (0.672)	Mean: 1.589 (0.396) Standard deviation: 0.913 (0.756)
-Log distance β	Mean: 2.495 (0.246) Standard deviation: 0.899 (0.178)	Mean: 2.827 (0.346) Standard deviation: 1.300 (0.399)
Differential preferences for city characteristics	Yes	Yes
City fixed effects	Yes	Yes
Log likelihood	-6,474.94	-6,478.24
Number of observations	59,415	59,415
Post-regression estimation of $\frac{\beta}{\alpha}$	1.466 (0.414)	1.779 (0.510)

Robust standard errors clustered by `city*home_province` 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 a more detailed explanation). 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>).

relax the unappealing assumption of IIA, these nested logit models do not change any of our major findings.

Next, in Table 6, we report regression results from mixed logit models. The two key parameters, α and β , are assumed to be jointly normal in column (1) and jointly log-normal in column (2). The log-normal assumption perhaps makes more sense because, based on economic theory, we expect both α and β to be positive. In both specifications, we assume that all other parameters are fixed. The estimated mean value of α is smaller in the log-normal specification (1.589 vs. 1.702); in contrast, the mean value of β is larger in the log-normal specification (2.827 vs. 2.495). As a result, the estimated ratio $\frac{\beta}{\alpha}$ (calculated from their mean values) is considerably larger in the log-normal specification (1.779 vs. 1.466). These estimates are qualitatively similar to what we obtained using conditional and nested logit models.³²

The results presented so far show that our baseline estimates are robust to alternative

³²Instead of a joint distribution, we also tried specifications in which α and β are assumed to be independently normal or log-normal. The estimated $\frac{\beta}{\alpha}$ are 1.533 and 1.497, respectively.

choice model specifications. As long as we allow for differential preferences for observed city characteristics, the estimated $\frac{\beta}{\alpha}$ is always around 1.5. Results from two different specifications both indicate that this elasticity is lower for male than female migrants.

To give these results some concrete meaning, we do the following calculation. Let's assume that we want to induce every migrant to move to the next farthest city in our sample.³³ For the median migrant in the sample, who is currently 225 km away from home, this supposed relocation implies a move of 68 km (or 30 percent) further away from home. Based on estimates in column (3) of Table 4, which allow for different elasticities between female and male migrants, we calculate the required income compensation. The results indicate that the monthly earnings for the median migrant need to increase by 655 yuan.³⁴ 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.

Our estimated income-distance elasticities seem to be very high. Naturally, we would like to check them against findings in the existing literature. Before we do so it is important to note that, to the best of our knowledge, we are the first to use this modeling framework to examine the income-distance tradeoff at the individual level. Existing studies typically regress migration flows on distance and income in destination locations, and then use the two coefficients to assess the income-distance tradeoff. Taking our method as the benchmark, it is easy to recognize that using aggregate data to measure this tradeoff has two problems: First, it does not take into account that the migrants may be very different from local residents at the destination, therefore the income level at the destination may not be their expected earnings after migration. Second, amenities are generally correlated with income level at the destination (Roback, 1982), and therefore when amenities are not properly controlled, the income coefficient may be biased upward or downward, resulting in an under- or over-estimated income-distance elasticity.

Keeping these limitations in mind, we nonetheless check what existing studies have found. In an early work on the economic approach to migration, Sjaastad (1962) reports that in the United States, "the attractiveness of a given destination was unaffected by a 10 percent gain in annual per capita labor earnings and a simultaneous 16 percent increase in distance," implying an income-distance elasticity of 0.63. Also studying U.S. internal migration, Davies et al. (2001) show that at 500 miles away from the origin, an increase of 45-63 1996 dollars in per capita income is needed to offset a one-mile increase in migration distance. If we assume the per capita income at destination to be \$24,300 (the 1996 U.S. per capita income), these results imply an income-distance elasticity between 0.93 and 1.30, substantially higher than the estimate by Sjaastad (1962).³⁵ Using data from Canada, Courchene (1970) finds the

³³For those who are already in the most distant city, we calculated the average extra distance for others and imposed that these individuals would move that much further away.

³⁴It is worth noting here that this back-of-envelope calculation assumes constant urban amenities. In practice, migrants who move further away from home villages almost always end up in cities with better amenities. Therefore, they do not demand a wage increase as sharp as suggested here.

³⁵This elasticity is even higher as one moves further away from the origin state. Davies et al. (2001)

income-distance elasticity to be in the range of 0.15-0.35, across different cohorts and model specifications.³⁶

Using data from Vietnam, Phan and Coxhead (2010) report an income-distance elasticity of 0.73. In China, findings in Fan (2005) suggest that this elasticity is between 0.352 and 1.054, based on different samples and model specifications. Also for China, results from Poncet (2006) imply an elasticity between 0.37 and 1.15, again across different samples and model specifications.

Therefore, we find that our estimates of the income-distance elasticity are considerably higher than those calculated from earlier studies. Since Fan (2005) and Poncet (2006) also study internal migration in China, we have given more thought to the difference between our estimates and their results. We note that neither Fan (2005) nor Poncet (2006) attempted to control for amenities in destination locations. In China, high-income cities, such as Shanghai, Guangzhou, and Hangzhou, generally also have high amenities, as suggested by the estimated city fixed effects in our regressions. Consequently, the estimated income coefficient is likely to be biased upward when the regression does not include amenity controls. We suspect that Fan (2005) and Poncet (2006) may have overestimated the income coefficient for this reason, which explains why their results give lower income-distance elasticities than ours.

5.3 Additional sensitivity analysis

We have performed some additional sensitivity analysis to assess the robustness of our findings. We estimate earnings in alternative ways; we also add more city level controls to the conditional logit regression. In each case, we examine how our baseline results are affected. As a benchmark, we focus on the baseline specification in column (3) of Table 4, in which we separately estimate the income-distance elasticity for female and male migrants. To ease comparison, the two elasticities from this baseline regression are reproduced in row (a) of Table 7.

For all regressions presented in Tables 4-6, we used earnings for each migrant in each destination city estimated from equation (13). In particular, to correct for potential selection biases, we used a second degree polynomial to approximate the unknown function of migration probability. Although it seems to be a reasonable choice, we want to explore whether this particular functional form affects our results. As an alternative, we now instead use a fourth degree polynomial of migration probability in the earnings equation. The estimated elasticities for female and male migrants are presented in row (b) of Table 7. Both elasticities are a little higher than in our baseline regression, but the differences are so small

estimate an empirical model using destination-to-origin income ratio (instead of two separate income variables) as an independent variable, so we cannot directly calculate the income-distance elasticity from their estimated coefficients. Our calculations here are based on the information in their Table 4 (p. 355).

³⁶Courchene (1970) tried many specifications. Our calculations are based on his results in Table V (p. 569) and equations 1 and 3 in Table VI (p. 570), where both distance and destination income enter the equation in log terms.

Table 7: Conditional logit regression results for sensitivity analysis

	Estimated $\frac{\beta}{\alpha}$ for FEMALE migrants	Estimated $\frac{\beta}{\alpha}$ for MALE migrants
(a) Baseline results from Table 4, Column (3):	1.571 (0.427)	1.454 (0.396)
(b) Correcting for selection biases using a quartic polynomial:	1.582 (0.420)	1.466 (0.390)
(c) Ignoring potential selection biases in income estimation:	1.631 (0.464)	1.510 (0.430)
(d) Controlling for differences in urban CPI and CPI growth (1988-2008) across destination provinces:	1.433 (0.372)	1.325 (0.344)
(e) Controlling for differences in secondary- and tertiary-sector employment shares across destination cities:	1.438 (0.365)	1.338 (0.339)
(f) Controlling for differences in mining, manufacturing, construction, wholesale and retail, and lodging and dining employment shares across destination cities:	1.209 (0.338)	1.147 (0.320)
(g) Controlling for number and share of migrants from the same home province to the destination city:	1.461 (0.391)	1.367 (0.366)
(h) Controlling for number and share of migrants from the same home province to the destination city, interacted with migration distance:	1.371 (0.365)	1.276 (0.340)

Robust standard errors clustered by city are in parentheses. In each specification, the interactions of individual and city characteristics are included to allow for differential preferences as in Table 4, column (3): 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, and per capita air pollutants), and therefore a total of 36 interactions.

that they are negligible.

We also try a specification that ignores potential selection biases, i.e., excludes the migration-probability terms from equation (13). The estimated elasticities under this specification are in row (c). Again, these elasticities are higher, but the differences are small. Therefore, the particular function form to deal with potential selection biases in our income regressions does not drive our main results. If anything, our baseline specification could have slightly underestimated the income-distance elasticities.

We next explore whether our main results are sensitive to the inclusion of more city level controls. Note that because we have the city fixed effects $(\bar{\theta}_1, \dots, \bar{\theta}_J)$ in our baseline regression, the average effect of any characteristic of city j on migrants' utilities would have been absorbed in $\bar{\theta}_j$. Therefore, city level controls could affect our estimation results only if migrants have heterogeneous preferences and different migrants attach different values to such city characteristics.

First, we consider the differences in cost of living across destination cities in our sample. Some careful research has documented that both local prices and their growth vary a great deal across different provinces in China (Brandt and Holz, 2006). For example, the consumption basket for a typical consumer could cost a lot more in the coastal city of Shanghai than the inland city of Chongqing. This difference could affect different migrants in different ways. For example, a married man may be more sensitive to high prices than a single man simply because the married man has a family to support. Consequently, other things equal, a married man may consider the low cost city of Chongqing more attractive than a single man, an effect not captured by $\bar{\theta}_j$ in our baseline regression.

To take this into account, we draw on the information about province-level urban CPI in China assembled and maintained by Brandt and Holz.³⁷ We construct two variables: one is the 2008 urban CPI at the destination province and the other the growth of urban CPI at the destination province during 1988-2008. We interact these variables with individual migrant characteristics (age, gender, education, and marital status) and add them to the conditional logit regression. The estimated income-distance elasticities for female and male migrants are presented in row (d) of Table 7. For both groups, the estimates are lower than the baseline results, but only moderately (by about 9 percent).

Next, we consider the differences in industry composition across destination cities. The idea is that different industries concentrate in different cities and some industries may be willing to pay more and thus more attractive to certain types of migrants. For example, a city with a larger share of construction jobs may be more attractive to male migrants; in contrast, a city with more manufacturing jobs may be more attractive to female migrants. Similarly, low status service jobs (such as those in the lodging and dining industry) may be

³⁷Brandt and Holz (2006) first reported their calculation of province-level CPI in China. They have since updated their calculation to the year 2010 and made it available at this website: <http://ihome.ust.hk/~socholz/SpatialDeflators.html>. For our purpose, similar data at the city level would be preferable, but we cannot find a reliable source for such data.

less attractive to better educated migrants.

We characterize a city's industry composition in two different ways. The first one is rather crude. Using information from the 2008 edition of the *Urban Statistical Yearbook of China*, we calculate the secondary- and tertiary-sector employment shares for each destination city. We then interact these variables with the four individual migrant characteristics and add them to the baseline regression. The estimated elasticities are presented in row (e). This again lowers our baseline estimates moderately, by about 8 percent.

Our second way to deal with industry composition measures employment shares at a more refined level. We first identify the five industries that are most likely to hire many rural-urban migrants: mining, construction, manufacturing, wholesale and retail, and lodging and dining. Again using information from the 2008 edition of the *Urban Statistical Yearbook of China*, we calculate the employment share of each of these five industries in each destination city. We interact each employment share with four individual migrant characteristics and add them to the regression. This turns out to have relatively large effects on our estimates. It reduces the elasticity from 1.571 to 1.209 for female migrants, and from 1.454 to 1.147 for male migrants. However, even these reduced estimates are quite large compared to those in the literature we discussed above.

The third set of controls concerns migration network. One might suspect that the willingness to pay for staying closer to the home village 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 bond 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 perhaps more difficult for a migrant to find a similar support network. For this reason, a migrant would appear willing to give up some income in order to stay close to home.³⁸

There is no lack of anecdotal evidence in support of this migration network hypothesis (Wei, 2008). To detect the network effect empirically, we obtained access to a large scale 1 percent population survey in China conducted in 2005, which is generally referred to as the "mini census." From a one-fifth random sample of this survey data,³⁹ we identify all the individuals that satisfy all of the following conditions: (i) they are at least 18 years old; (ii) they have a rural *hukou*; (iii) they do not have a local *hukou*; (iv) they have left their *hukou* place for at least half a year; and (v) the reason for migration is to find work or do business. This gives us a sample of 116,710 migrants who originated in rural areas in China, which allows us to measure the flows of rural migrants across regions.

For each migrant in our RUMiC sample, we identify his or her home province. We then count the number of rural migrants in the mini-census data who have moved from that

³⁸Of 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.

³⁹For some reason, the government agency decided to release only one-fifth of the data to the research community. As far as we know, no researchers have access to more than this one-fifth subsample of the data.

province to each of the 15 destination cities in our RUMiC sample; we also calculate the share of migrants in that home province who have moved to each of the 15 destination cities.⁴⁰ We use these two variables to measure a migrant’s network size in each destination city. These migration network variables are added to the logit regression in two different ways. First, we include them directly in the utility function, just like the log distance variable. The estimated elasticities are presented in row (g) of Table 7. Again, these controls lower the estimates, but only slightly. Second, instead of adding the network variables directly to the utility function, we interact each of them with log distance in the regression. In this case, we assume that large migration networks per se are not valued but they help reduce the disutility of migration distance. The estimated elasticities are in row (h). They are smaller than those from the first specification, but of similar order.

Overall, these sensitivity analyses with more controls suggest that our baseline results are rather robust; they cannot be easily explained away. Consider the most conservative estimates in row (f) of Table 7. The income distance elasticity is 1.209 for female migrants and 1.147 for male migrants. These parameters imply that other things equal, one has to pay an extra 495 yuan a month to induce the median migrant to move 68 km further away from home. Clearly, this is still a very steep tradeoff.

5.4 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 for closer destinations 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. 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 to put this into perspective. Consider a trip from Wuhan to Guangzhou by express train. The total distance is 1,069 km and the price for an economy-class ticket is only 140 yuan.⁴¹ That is, on average it only costs 8.91 yuan to travel 68 km, which is totally insignificant compared to the 655 yuan a month the median migrant is willing to pay to avoid such travel.

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 alone cannot explain why long-distance moves are generally associated with

⁴⁰Our data actually allows us to calculate prefecture-to-city migration flows. However, such numbers are small and necessarily contain a larger portion of noise. In addition, anecdotal evidence suggests that when people move outside of their home province they indeed tend to form a close bond with all fellow migrants from the same province. For these reasons, we aggregated the data at the home province level instead of the origin prefecture level.

⁴¹Source of this information: <http://open.baidu.com/train/search>. It is the ticket price as of March 10, 2011; the price might have been even lower in 2008 when our survey data were collected.

higher earnings as found in the data. Moreover, if localized job opening 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 production to those regions. And lastly, 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, because education tends to increase a person's capability of obtaining and utilizing information. 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.

One last possible reason behind this high income-distance elasticity is the particular institutional context of rural-urban migration in China. On the one hand, under the household responsibility system, all these migrant workers with rural *hukou* have access to some farm land in their home villages, which is typically allotted to each family based on headcount. When they migrate to urban areas, 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. On the other hand, in destination cities, the *hukou* system prevents rural migrants from assimilating into local urban communities. Because rural migrants do not have urban *hukou*, their school-aged children either are not allowed to enter urban public schools or have to pay some extra fees to go to such schools. Similarly, their senior parents are not eligible for subsidized health care in cities available to regular urban residents. In addition, many urban sector jobs, such as those in government, finance, and large state-owned enterprises, are reserved for workers with urban *hukou* and thus beyond the reach of rural migrants. Since these migrants have land in their home villages and are treated as outsiders in cities, they tend to consider their move to urban areas temporary. They think of their home villages as a fallback in the event jobs in urban sectors are not easily available (Yang, 1997; de la Rupelle et al., 2009). For this reason, many leave their parents and children behind in home villages, and may decide to stay close in case of emergency.

In some additional exploratory analysis, we experiment by adding a few more individual characteristics to the specification in column (3) of Table 4, all interacted with log migration distance. We find that the income-distance elasticity tends to be bigger for migrants who have at least one parent in poor health and that the elasticity tends to be smaller for migrants with more siblings. Both results seem to be consistent with the notion that parents are left behind in the village and migrants may have to go back to take care of them when needed. We also find that when the spouse of the migrant stays behind in the home village, perhaps to take care of children, parents, or family properties, the income-distance elasticity is bigger, which makes perfect sense. Whereas these results are suggestive evidence that family members and properties left in the home village are related to the income-distance elasticity, we are unable to assess how important this factor is compared to other factors such as the value of local networks.

It is clearly important for future work to pin down the key determining factors behind the high income-distance elasticity. If it is mainly a result of government policy distortions, then a reform aimed at reducing this elasticity can potentially increase labor mobility in China and further enhance the country's advantage in labor-intensive industries. On the other hand, further study may reveal that this high income-distance elasticity is not something that is easily manipulated. In that case, our findings here suggest that the spacial dimension of China's urbanization policies can have important welfare implications. For example, if two cities are identical in every other aspects except that one is closer to surplus labor in rural areas than the other, then it makes economic sense to allow the closer city to absorb more rural-urban migrants. Government policy should facilitate such developments instead of distorting the process by encouraging migration flows toward the more distant city.

6 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 are willing to give up in order to stay closer to home villages.

We build a simple model in which migrants from rural areas choose among potential destination cities to maximize utility. The distance between a migrant's home village and destination city is assumed to enter the utility function directly. Using 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 baseline estimates imply an income-distance elasticity of 1.5: To induce an individual to migrate 10 percent further away from home, the wage paid to this migrant has to increase by 15 percent. We tried different model specifications and experimented with a wide variety of controls. Even our most conservative estimates suggest an income-distance elasticity close to 1.2. This elasticity appears to be very high. It varies very little with migration distance; it is slightly higher for female 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 villages. We suspect that it has to do with the specific social and institutional context in China. Pinning down the exact reasons behind this steep 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] Banerjee, Abhijit, Esther Duflo, and Nancy Qian (2012). “On the Road: Access to Transportation Infrastructure and Economic Growth in China,” NBER Working Paper 17897, National Bureau of Economic Research, Cambridge, MA.
- [2] Bao, Shuming, Orn B. Bodvarsson, Jack W. Hou, and Yaohui Zhao (2009). “Migration in China from 1985 to 2000,” *Chinese Economy* 42, 7–28.
- [3] Baum-Snow, Nathaniel, Loren Brandt, Vernon Henderson, Matthew Turner, and Qinghua Zhang (2012). “Roads, Railroads and Decentralization of Chinese Cities,” unpublished manuscript, Department of Economics, Brown University.
- [4] Baum-Snow, Nathaniel and Matthew Turner (2012). “Transportation and the Decentralization of Chinese Cities,” unpublished manuscript, Department of Economics, Brown University.
- [5] 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.
- [6] 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.
- [7] Brandt, Loren and Carsten A. Holz (2006). “Spatial Price Differences in China: Estimates and Implications,” *Economic Development and Cultural Change* 55, 43–86.
- [8] Chan, Kam Wing and Li Zhang (1999). “The Hukou System and Rural-Urban Migration in China: Processes and Changes,” *China Quarterly* 160, 818–855.
- [9] Courchene, Thomas J. (1970). “Interprovincial Migration and Economic Adjustment,” *Canadian Journal of Economics* 3, 550–576.
- [10] Dahl, Gordon (2002). “Mobility and the Return to Education: Testing a Roy Model with Multiple Markets,” *Econometrica* 70, 2367–2420.
- [11] 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.
- [12] 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.
- [13] Falaris, Evangelos M. (1987). “A Nested Logit Migration Model with Selectivity,” *International Economic Review* 28, 429–443.

- [14] Fan, C. Cindy (2005). “Modeling Interprovincial Migration in China, 1985–2000,” *Eurasian Geography and Economics* 46, 165–184.
- [15] Hahn, Jinyong and Geert Ridder (2010). “The Asymptotic Variance of Semi-parametric Estimators with Generated Regressors,” unpublished manuscript, Department of Economics, UCLA.
- [16] Kennan, John and James R. Walker (2011). “The Effect of Expected Income on Individual Migration Decisions,” *Econometrica* 79, 211–251.
- [17] Kong, Sherry Tao (2010). “Rural–Urban Migration in China: Survey Design and Implementation,” in Xin Meng and Chris Manning with Li Shi and Tadjuddin Noer Efendi (eds.), *The Great Migration: Rural–Urban Migration in China and Indonesia*, pp. 135–150, Cheltenham: Edward Elgar.
- [18] Lee, Lung-Fei (1983). “Generalized Econometric Models with Selectivity,” *Econometrica* 51, 507–512.
- [19] Lin, Justin Y., Gewei Wang, and Yaohui Zhao (2004). “Regional Inequality and Labor Transfers in China,” *Economic Development and Cultural Change* 52, 587–603.
- [20] McFadden, Daniel (1974). “Conditional Logit Analysis of Qualitative Choice Behavior,” in P. Zarembka, ed., *Frontiers in Econometrics*, Academic Press, New York, 105–142.
- [21] 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.
- [22] McFadden, Daniel and Kenneth Train (2000). “Mixed MNL Models for Discrete Response,” *Journal of Applied Econometrics* 15, 447–470.
- [23] Murphy, Kevin M. and Robert H. Topel (1985). “Estimation and Inference in Two-Step Econometric Models,” *Journal of Business and Economic Statistics* 3, 370–379.
- [24] Nakosteen, Robert A. and Michael Zimmer (1980). “Migration and Income: The Question of Self-Selection,” *Southern Economic Journal* 46, 840–851.
- [25] Petrin, Amil and Kenneth Train (2010). “A Control Function Approach to Endogeneity in Consumer Choice Models,” *Journal of Marketing Research* 47, 3–13.
- [26] Phan, Diep and Ian Coxhead (2010). “Inter-Provincial Migration and Inequality During Vietnam’s Transition,” *Journal of Development Economics* 91, 100–112.
- [27] Poncet, Sara (2006). “Provincial Migration Dynamics in China: Borders, Costs and Economic Motivations,” *Regional Science and Urban Economics* 36, 385–398.

- [28] Roback, Jennifer (1982). “Wages, Rents, and the Quality of Life,” *Journal of Political Economy* 90, 1257–1278.
- [29] Robinson, Chris and Nigel Tomes (1982). “Self-Selection and Interprovincial Migration in Canada,” *Canadian Journal of Economics* 15, 474–502.
- [30] Schwartz, Aba (1973). “Interpreting the Effect of Distance on Migration,” *Journal of Political Economy* 81, 1153–1169.
- [31] Sinnott, R. W. (1984). “Virtues of the Haversine,” *Sky and Telescope* 68, 159.
- [32] Sjaastad, Larry A. (1962). “The Costs and Returns of Human Migration,” *Journal of Political Economy* 70, No. 5, Part 2, 80–93.
- [33] 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.
- [34] Train, Kenneth (2009). *Discrete Choice Method with Simulation*, 2nd edition, Cambridge University Press, Cambridge, MA.
- [35] Vanderkamp, John (1971). “Migration Flows, Their Determinants and the Effects of Return Migration,” *Journal of Political Economy* 79, 1012-1031.
- [36] Wei, Cheng (2008). *An Investigation of Farmers Turned Workers in China*, Beijing: Law Press.
- [37] Yang, Dennis Tao (1997). “China’s land arrangements and rural labor mobility,” *China Economic Review* 8, 101–115.
- [38] Zhang, Junfu and Zhong Zhao (2012). “Social-Family Network and Self-Employment: Evidence from Temporary Rural-Urban Migrants in China,” unpublished manuscript, Department of Economics, Clark University.