

# Destination Choices of Chinese Rural-Urban Migrant Workers: Jobs, Amenities and Local Spillovers\*

Zhiling Wang<sup>1,2</sup> and Lu Chen<sup>3</sup>

<sup>1</sup>Erasmus School of Economics, Erasmus University Rotterdam, the Netherlands

<sup>2</sup>Tinbergen Institute, the Netherlands

<sup>3</sup>School of Finance, Nankai University, China

April 5, 2019

\*Corresponding author: Zhiling Wang. Email: [z.wang.vu@gmail.com](mailto:z.wang.vu@gmail.com). We are grateful to Daniel Broxterman and other guest editors of the Journal of Regional Science special issue on “endogenous amenities and cities” for editorial guidance, Chun Kuang for insightful discussant comments, and an anonymous referee for constructive suggestions. This paper has also benefited from comments and suggestions from Thomas de Graaff, Paul Elhorst, Shuaizhang Feng, Yuming Fu, Michiel Gerritse, John Stillwell, Frank van Oort, Chunbing Xing, Sen Xue, Yihua Yu, Siqi Zheng, and the participants of seminars at Jinan University, Guangzhou; Wageningen University and Research, Wageningen; the 2016 Floating Population Forum, Beijing; RUSE conference, Xiamen; “sustainable urban and regional development” conference at Renmin University, Beijing; ERS conference, Groningen; and the “endogenous amenities and cities” symposium at Florida State University, Tallahassee. Lu Chen thanks the National Planning Office of Philosophy and Social Science, Grant Number 16BRK019 for the financial support. We thank the research assistants Ling Cao, Jinxu Wang, Kangkai Wang, and Bingfei Zhai for their contributions to this project. All remaining errors are ours.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi:10.1111/jors.12444

## Abstract

Using the 2014 China Migrants Dynamic Survey, we analyze rural-urban migrant workers' destination choices after the global financial crisis, with an emphasis on jobs, amenities, and local spillovers. By using an equilibrium-sorting model, this paper disentangles local spillovers from local attributes in the estimation process. We employ both an artificial instrumental variable and the provincial highway passenger flow in 1979 to tackle the endogeneity issue. After controlling for the network effects of migrants from the same origin, we found a separate and strong preference for co-locating with a large population of migrants, regardless of origin. The results remain robust when we take into account labor supply-driven migration, spatial autocorrelation between provinces, different industry definitions, and regional differences within provinces. Our results imply that due to institutional barriers, the rural migrant community will still be a very important factor in the foreseeable future. In addition, as the ongoing industrial upgrading and transfer policies in China may lead to a westward movement of rural-urban migrants, the movement will be expedited when the older, less-educated, or lower-income migrants relocate.

**Keywords:** Rural-urban migration in China, destination choice, labor market, amenities, local spillovers

**JEL codes:** J24, J61, R23

# 1 Introduction

China's rise as the "world's factory" and its unprecedented economic growth are closely intertwined with epic-scale urbanization. The number of rural migrants working in Chinese cities grew from approximately 39 million in 1997 to 172 million in 2017. This large-scale movement of labor is viewed as playing a critical role in the long-term economic growth of Chinese cities (Chan, 2012; Combes et al., 2015; Meng et al., 2013). However, a potential challenge for sustaining growth might lie in the degree to which the flow of rural-to-urban workers is comprised of temporary rather than permanent migrants.

Labor mobility in China is restricted by the system of household registration, known as hukou. Established in the 1950s, this system records a household's official place of residence and classifies them as either agricultural (rural) or non-agricultural (urban).<sup>1</sup> On the basis of this classification, the system determines eligibility for and levels of important welfare benefits, such as pensions, housing, medical care, and education. Benefits for urban hukou holders are more generous than rural, and benefits can only be received in the home province where the hukou is registered, as discussed more fully in the next section. Unless this differential treatment is changed, many rural migrants and their children will likely intend to eventually return to their hometowns. In this way, rural-urban migrants in China differ from the migrants typically analyzed in studies of urbanization in developing countries. We conjecture that when rural hukou holders make location decisions, they focus mostly on immediate economic opportunities and network support in the face of this systemic discrimination, and very little on other factors such as amenities that might weigh more heavily when considering a permanent relocation.

Since the reform of China's economy in 1978, coastal provinces in the eastern region have received large migrant inflows drawn from rural areas in the middle and western regions, which have lagged behind economically.<sup>2</sup> China's entry into the World Trade

---

<sup>1</sup>People inherit the hukou status at birth, so that a child born in Shanghai to rural migrants from a western province would have the same rural hukou as her parents. Converting from a rural to an urban hukou status can be extremely difficult, except for special reasons, such as attending universities, being displaced after land expropriation and serving in the army (Chan and Buckingham, 2008; Hu et al., 2011).

<sup>2</sup>We adopt the classification of regions used in the Rural Workers Monitoring Survey, National Bureau of Statistics of China. The eastern region includes the provinces of Beijing, Tianjin, Hebei, Liaoning,

Organization in 2001 intensified demand for the labor-intensive and export-oriented manufacturing industries located in the eastern region, and surplus rural labor migrated to cities there in search for better economic opportunities. Although the coastal provinces suffered the most from the Global Financial Crisis (GFC) in late 2008, due to substantial reduction in the global demand for imported goods from China, the GFC did not create long-term unemployment for migrant workers, and most rural migrant workers returned to cities in 2009 as more jobs became available.<sup>3</sup>

Presumably, eastern coastal provinces are still expected to be attractive to rural-urban migrant workers after the GFC. However, based on surveys of the National Bureau of Statistics, we find interesting changes in the spatial patterns of migration in the years following the GFC. As shown in Panel (A) of Figure 1, the proportion of rural workers migrating to the eastern region shows a continuous decline of approximately 8 percent. In Panel (B) of Figure 1, we show that rural-urban migration is increasingly an intra-province phenomenon, and the rate of growth is higher in the middle and western regions. In short, workers are migrating more to nearby destinations after the GFC. That is, the migration flow is changing from a predominantly inland-coast direction to both inland-inland and coast-coast.

[FIGURE 1 ABOUT HERE]

What is causing the spatial redistribution of rural-urban migration flows? Studying the new migration pattern is of great importance, as migration is closely related to the spatial allocation of human capital and to local economic development. Understanding rural-urban migrant sorting patterns is also helpful to local migrant population service centers, enabling them to design more efficient policies.<sup>4</sup> This paper sets out to understand what factors are driving the phenomenon of westward and short-distance

---

Shanghai, Jiangsu, Zhejiang, Fujian, Shandong, Guangdong, and Hainan. The middle region includes the provinces of Shanxi, Jilin, Heilongjiang, Anhui, Jiangxi, Henan, Hubei, and Hunan. The western region includes the provinces of Inner Mongolia, Guangxi, Chongqing, Sichuan, Guizhou, Yunnan, Tibet, Shaanxi, Gansu, Qinghai, Ningxia, and Xinjiang.

<sup>3</sup>According to Chan (2010) and Kong et al. (2010), the GFC caused approximately 10 percent of total migrants to return to rural villages at the end of 2008 and most of them temporarily went back to the rural farming sector as an employment buffer. Those who still worked in cities the next year, suffered very little in terms of employment, working hours and earnings.

<sup>4</sup>In recent years, the central government began setting up many local branches specifically aimed at helping migrants in various aspects of life, such as settlement and social integration. See <http://www.ldrk.org.cn/> for more details.

migration. We study the destination choice of rural-urban migrant workers in China, with an emphasis on comparing the roles of jobs, amenities and, in particular, local spillover effects.

The local spillovers in this paper refer to the channels through which the presence of a large rural migrant inflow might affect location choice separately from the presence of migrants from one's own province. We conjecture that the presence of a large rural migrant inflow is a favorable local attribute for future potential migrants, because it facilitates labor market pooling and serves to limit the effects of various forms of discrimination faced by rural hukou holders in cities. Although a large migrant population is likely associated with a deterioration in housing quality, because rural migrants are competing for limited cheap housing units on the rental market, the costs might be smaller than the aforementioned benefits.<sup>5</sup>

We analyze the destination choices of 58,595 rural-urban migrant workers, using the 2014 China Migrants Dynamics Survey. A retrospective question on the timing of migration allows us to study the sorting patterns of migrants between 2008 and 2014, the critical transition period after the GFC for China. Our main conclusions are as follows. For province characteristics, the pull effects of jobs and local spillovers are more important than those of amenities. Employment in secondary industries contributes substantially to province's attractiveness for potential migrant workers, whereas the presence of a large rural migrant inflow increases the province's attractiveness by a similar magnitude. Good medical resources turn out to be attractive for rural migrants, but we do not find such a pull effect for educational resources. Only higher-income workers place greater weights on medical and educational resources when choosing destinations. For individual heterogeneous preferences, the preference for co-location with other migrants, regardless of origin, is especially strong among the older, less-educated and lower-income groups. This might reflect the migrants' high demand for social support from peers. Our results imply that perhaps, due to institutional barriers faced by rural hukou holders, the presence of a large migrant community will likely be a very important factor affecting location choice for the foreseeable future. In addition, the trend in China has been for manufacturing firms to relocate more labor-

---

<sup>5</sup>The situation is somewhat similar to international migration in developed countries. International migrants usually choose to live where there are others with the same ethnic background and a presence of other immigrant groups as well. Co-ethnic networks provides assistance in job information and initial settlement, whereas a high share of foreign-born population reflects that the area is possibly a more tolerant environment to foreigners (see, e.g., Åslund, 2005; Buckley, 1996; Zavodny, 1999).

intensive plants to the middle and western regions in response to ongoing industrial upgrading and transfer policies. Our results indicate that the flow of rural migrants to these new employment opportunities will likely accelerate as spillover effects draw in still more rural migrants, in particular from the older, less-educated, or lower-income demographic groups.

We contribute to the literature on urbanization in China in two aspects. First, we examine new westward and short-distance migration pattern after the GFC by using a large up-to-date micro dataset of migrant workers. We study the relative importance of jobs, amenities, and local spillovers among rural migrants on location choice. This study helps to better understand the determinants of rural-urban migration, and to predict what future demographic patterns will be in response to changing urbanization or industrial policies in China. Second, we pay special attention to a critical but often neglected determinant of destination choice, local spillovers. In the previous literature, by simply observing people choosing to move to a specific place on a large scale, it is not clear whether the attractiveness of the destination can be attributed to certain location-specific features or to agglomeration effects and other externalities. Controlling only for location-specific fixed effects would mix these effects. Other papers examining migrant location choices in China have often used province or city level fixed effects (see, e.g. Liu and Xu, 2017; Xia and Lu, 2015; Zhang and Zhao, 2013) in a discrete choice framework. By using the equilibrium-sorting framework developed by Bayer and Timmins (2007), we are able to test for the impacts on location choice of various observed local attributes and recover estimates of local spillovers in the presence of unobserved characteristics.

The remainder of the paper is organized as follows. Section 2 develops the main factors examined in this paper. In section 3, we explain the sorting model used for our empirical analyses and our identification strategy. Section 4 describes the dataset and the variables, and then, section 5 presents and interprets the estimation results. Section 6 gives the results of different robustness checks, all of which are found to be in line with our main findings. The final section discusses potential policy implications and avenues for future research.

## 2 Jobs, Amenities, and Local Spillovers

This section reviews related studies, and presents our hypotheses with regard to jobs, amenities, and local spillovers. There are two types of studies on the migration decisions of rural-urban migrants in China. Earlier studies of rural-urban migration primarily focused on the individual decision to become a migrant (see, e.g., Liang and White, 1997; Wang and Zuo, 1999; Zhao, 1999; Zhao, 2005). The initiation of a migration decision is closely related to one's farmland size, rural enterprise development, local rural income, number of early migrants from the village, and so forth.

Several recent studies (Su et al., 2018; Xing and Zhang, 2017; Zhang and Zhao, 2013) focus on a subsequent decision related to migration, i.e., how rural migrants choose their destinations, and what their sorting patterns are. Economic improvement is usually the most important incentive, which is found to be a positive factor impacting the rural migrant's destination choice in all previous studies. Migrants tend to emigrate to areas with relatively high wages and low moving costs. According to Lin et al. (2004), China's widening income disparity between its coast and inland regions, as well as urban and rural regions, is consistent with increased responsiveness of migration to regional income differentials. In more developed areas, there are more job opportunities as well as higher income levels. This pull effect has also been empirically tested for the entire population of migrant workers in China (Fu and Gabriel, 2012; Xia and Lu, 2015), but not specifically the rural-urban type.<sup>6</sup> Similarly, we conjecture that jobs and wages are playing an unambiguously positive role to attract more rural migrants to cities.

In this paper, amenities refer specifically to public goods. It is worth mentioning that the central government in China has initiated reforms aiming to increase migrant participation in urban employee health insurance and pension programs in the 2008 Labor Contract Law (Giles et al., 2018). Although the percentage of rural migrant workers enrolled in urban health insurance or pension programs is still low (30 percent in 2015), access to urban social insurance programs is open to migrant workers. In contrast, the situation is worse for educational resources. The right to attend urban public schools in most cities is still associated with hukou status. Public schools have limited quotas. Most migrant children in cities can only go to a special type of migrant school, where the teachers and managers usually come from rural areas also. These

---

<sup>6</sup>In the context of international migration, we refer to Wang et al. (2018) for an exhaustive literature review of determinants for migrants' location choice.



migrant schools began to emerge in the 1990s and quickly became the major venue for the education of migrant children (Chen and Feng, 2017; Lai et al., 2014). As medical resources and educational resources are two of the most important public goods, we conjecture that a typical migrant is likely to be concerned with them.<sup>7</sup>

Networks play significant roles in location choice. The previous literature has focused extensively on network effects of migrants from the same origin (see, e.g. Liang et al., 2017; Rozelle et al., 1999; Zhao, 2003), but has ignored network effects of other migrants from different origins. The central feature here is that the payoffs from choosing a location partly depend on the number of other rural migrants choosing the same location in equilibrium, independent of their origins. Therefore, our key variable of interest, i.e., local spillovers, is measured by the share of individuals choosing a specific location among all rural migrants. In the following three paragraphs, we explain possible channels where the presence of a large rural migrant inflow might affect location choice.

As the system of household registration in China causes the migrant population to be segregated from the urban host population, rural migrant workers are treated differently from their urban counterparts in terms of occupational attainment and wages, and most of them, indeed, experience some form of the discrimination (Démurger et al., 2009; Kuang and Liu, 2012; Meng and Zhang, 2001). In fact, the rural migrants' self-identities are at least two-dimensional: (i) the perception of the culture and language of one's place of origin and (ii) the perception of being a rural migrant in cities under segregated economies of rural-urban division. On one hand, social networks formed by migrants from the same place ("Lao Xiang" in Chinese) emerge as effective mechanisms in facilitating rural-urban migration in China (Zhao, 2003). On the other hand, migrants from other places of origin can also provide important networks to make the groups of rural migrants less marginalized. "Nong Min Gong" is what the whole population of rural-urban migrants is called. The non-local identity limits their free access to public services in cities and to some extent, hurts their feelings (Chan, 2012; Chen, 2005). Hence, the presence of more rural migrants in the same destination would serve

---

<sup>7</sup>We also tried to include housing price as one of the important local attributes in an earlier version of this paper. We will discuss in detail in later paragraphs that migrant workers are only competing for low-cost dwelling units in cities. The average housing price publicly available cannot reflect the bottom portion of the rental market. As was shown in our earlier results, the marginal contribution to province's attractiveness is insignificant.



as a large community in which they might form social capital.<sup>8</sup> Social linkages and networks are more likely to be built among rural migrants themselves than to be built between rural migrants and urban citizens, due to their similar backgrounds. These social interactions provide a pool of information about settlement, jobs, and many other aspects of life, specific to the rural hukou holders.

The way more rural migrants choosing the same destination makes it attractive to still more can also be attributed to labor market reasons. As rural migrants sort into a destination, the density of the labor market increases from the supply side. Simultaneously, firms locate in destinations with large migrant inflows and thus increases a province's attractiveness for future rural migrants. Marshallian externalities, referring to the concentration of same-industry production, contribute substantially to China's industrial agglomeration (Lu and Tao, 2009). For example, the Pearl River Delta economic zone developed its industries in clusters after the economic reform in China. It has now become a major global manufacturing base. The specialized agglomeration of firms and workers generates potential benefits through labor market pooling. Workers gain industry-specific skills via working, self-learning, and vocational training. There is a reliable demand for specialized workers. When they change jobs, they do not have to change their skill set or move to a new destination, so they have lower search costs and quicker matching processes.

Nevertheless, a negative effect of many rural migrants in the same destination could possibly arise due to the limited housing in cities. The suppliers of urban housing have overlooked the needs of the migrant population, despite its considerable size. As rural migrant workers are excluded from the housing-distribution system, they rent low-cost dwelling units or live in enterprise dormitories (Song et al., 2008; Wu, 2004). Those affordable dwelling units are usually found in "villages in the city" ("Cheng Zhong Cun" in Chinese), located in both the outskirts and downtown segments of cities. Local farmers living in villages are allowed to construct housing units and rent them out. Many of these housing units are equipped with very poor facilities and situated in neighborhoods with inadequate infrastructure. The other source of housing, enterprise dormitories, are commonly present in construction and manufacturing sites. Dormitory-style housing provides smaller usable areas, making these housing options

---

<sup>8</sup>The reference group for a given population is based on spatial proximity and other dimensions. For Chinese rural-urban migrant workers confronting different types of populations, they may refer to non-migrants in the same village, a cohort from the same clan culture, natives in cities for comparison or decision-making on other things (Akay et al., 2012).

difficult for families to live in. Based on the latest survey of Rural Urban Migration in China<sup>9</sup> approximately 50 percent of rural migrants report that their employers provide accommodation, implying that the rest of the housing demand must be met in the rental market, most likely in these “villages in the city”. Migrant workers are quite constrained in their housing budgets. According to Zheng et al. (2009), on average, migrants are unwilling to spend more than 19 percent of their total income on housing. They prefer small dwelling units to save more money, even though the housing units are overcrowded and located in poorly served neighborhoods. In that case, more migrants sorting into one province means a continuing deterioration of living quality and thus a decrease in destination’s attractiveness for future rural migrants.

To sum up with regard to local spillovers, we conjecture that the presence of a large rural migrant inflow is a favorable local attribute for future potential migrants, because it facilitates labor market pooling and serves as another large community to deal with different forms of discrimination in cities. Although it is likely to cause a deterioration of housing quality because rural migrants are competing for limited cheap housing units on the rental market, whether the cost might be smaller than the aforementioned benefits is an empirical question we seek to answer.

### 3 Empirical Model

This section describes a sorting model for migration behaviors that can be used to assess the relative importance of jobs, amenities and local spillovers in destination choice. It is a widely accepted assumption that one’s locational preferences are expressed by physical migration, which is often called “voting by foot” (Tiebout, 1956). The equilibrium-sorting model proposed by Bayer and Timmins (2007) (and a slightly different version in Bayer and Timmins (2005) and Bayer et al. (2004)) inherits this classic assumption and is the starting point for our methodological framework. We use it to examine migrant workers’ heterogeneous preferences and the contribution of different province attributes to a location’s attractiveness. Our empirical analyses rely on the differentiated-product discrete-choice approach, typically referred to as the Berry-Levinsohn-Pakes (BLP) method in the literature (Berry et al., 1995). In addition

---

<sup>9</sup>The survey was initiated by the Australian National University. We calculate the percentage for new respondents who entered the survey in 2016.

to Bayer and Timmins (2007), the BLP method has been applied recently in a number of empirical studies of location choice (see, e.g. Levkovich and Rouwendal, 2014; Van Duijn and Rouwendal, 2013; Wang et al., 2016).

This model allows us to measure the size of local spillovers based on the location decisions of migrant workers in the presence of unobservable local attributes, employing the internal logic of the sorting model itself. The estimation process distinguishes the contribution of individual preferences from regional factors (including the unobserved attributes) in two steps.<sup>10</sup> Although the model has been developed to study the distribution of a whole population over a number of localities, it can also be applied to a specific group of individuals (Bayer and Timmins, 2007). In that case, local spillovers refer to self-segregating preferences among the individuals in that group. In this paper, we are focusing on the rural-urban migrant population in particular, and their sorting behaviors should be seen independently from the urban-urban migrant population. The latter group faces much fewer institutional barriers, and they differ substantially from the rural-urban migrant population in terms of educational level, income level, profession and migration behavior.

The dataset is a representative survey of migrant workers in China, with approximately 85 percent being rural-urban migrant workers. Our sample provides good spatial coverage with 58,595 observations and 26 provinces. The substantial variation across the locations that form the consideration set for each origin makes identification possible. One limitation of using this methodology in the context of China is that the individuals are a selected group of rural people who have decided to leave rural areas and move to cities. This means that this group might be more opportunity-seeking and may be more responsive to province attributes in making destination choices. Therefore, all coefficients should be interpreted as lower bounds relative to a less selected population.<sup>11</sup>

---

<sup>10</sup>In previous studies, only Xing and Zhang (2017) employs a similar idea in modeling the rural-urban migrants' location choice, whereas the focus is not on local spillovers among rural migrants themselves.

<sup>11</sup>This limitation applies to all discrete choice models on destination choice. The selection problem can be weakened to a certain extent if data about the migrants' counterpart group in the rural areas are available. Unfortunately, we do not have this information to look further into the differences between movers and stayers in rural areas.

### 3.1 Equilibrium-sorting model

We present the utility-maximizing location choice model in the context of rural to urban migration in China. A population of migrants, indexed by  $m = 1, \dots, M$  is modeled. The migrants' province origins are indexed by  $i = 1, \dots, I$ , and they choose a destination province from all possible alternatives, indexed by  $j = 1, \dots, J$ . Each migrant maximizes an indirect utility function in choosing destinations as follows:

$$\max_j V_{mj} = \mathbf{Z}_j' \beta_m + \alpha_m \sigma_j + \mathbf{P}_{ij}' \gamma + \eta_j + \varepsilon_{mj}. \quad (1)$$

$V_{mj}$  describes the utility that migrant  $m$  in origin province  $i$  derives from living in province  $j$ . Each province  $j$  is described by the following:

- $\mathbf{Z}_j$  an observable vector of province attributes;
- $\sigma_j$  the share of individuals who choose this location among all survey respondents;
- $\mathbf{P}_{ij}$  province pair-specific information for each origin  $i$  and destination  $j$  combination;
- $\eta_j$  province-specific unobservable characteristics.

Estimation requires that migrants do not sort across locations based on the unobserved qualities of the migrants after taking into account the common component captured by the location fixed effect  $\eta_j$ .

The taste parameters  $\beta_m$  and  $\alpha_m$  in equation (1) indicate that the individual preference for a particular province characteristic is not the same for every migrant, but is interacted with individual heterogeneity. It is a function of the migrants' individual characteristics,  $\mathbf{X}_m$ , such as age and gender, and is written as follows:

$$\beta_m = \beta_0 + \mathbf{X}_m' \beta_1, \quad (2)$$

$$\alpha_m = \alpha_0 + \mathbf{X}_m' \alpha_1. \quad (3)$$

We can rewrite our indirect utility function by substituting equations (2) and (3) into

equation (1), as follows:

$$V_{mij} = \lambda_j + \mathbf{Z}'_j \mathbf{X}_m \beta_1 + \sigma_j \mathbf{X}_m \alpha_1 + \mathbf{P}'_{ij} \gamma + \varepsilon_{mj}, \quad (4)$$

where

$$\lambda_j = \mathbf{Z}'_j \beta_0 + \alpha_0 \sigma_j + \eta_j. \quad (5)$$

### 3.2 Identification

The model is estimated using a two-stage procedure due to Berry et al. (1995).

*In the first stage*, equation (4) is estimated using multinomial logistic regression with alternative-specific constants. The coefficients  $\lambda$ ,  $\beta_1$ ,  $\alpha_1$  and  $\gamma$  are obtained:

- $\lambda$  measures province attractiveness;
- $\beta_1$  describes the individual heterogeneous preferences for various province attributes;
- $\alpha_1$  describes the individual heterogeneous preferences for other migrants choosing the same province;
- $\gamma$  describes the gravity force between origin  $i$  and destination  $j$ , such as the migrants' network from the same place of origin.

*In the second stage*, equation (5) is estimated using ordinary least squares or instrumental variables. The estimated  $\lambda$  values are regressed on  $\mathbf{Z}$  and  $\sigma$ , to obtain the coefficients  $\beta_0$  and  $\alpha_0$ :

- $\beta_0$  describes the contribution of observed province attributes to province's attractiveness;
- $\alpha_0$  describes the contribution of local spillovers to province's attractiveness.

We have to address potential endogeneity in the second step, as the unobserved province characteristics ( $\eta$ ) are very likely to be correlated with both the regressors and

province's attractiveness ( $\lambda$ ), which may cause estimation bias. Therefore, we resort to instrumental variable (IV) estimation. The most likely endogenous variable is  $\sigma$ , which measures the share of migrant workers in each province in the sample of migrant workers. If, for whatever reason (prospects for future development, productivity shocks, etc), many migrant workers choose province  $j$ , the coefficient for the variable of interest will be biased upwards. The unobserved characteristics  $\eta$  then leads to an overestimation of the coefficient  $\alpha_0$ . The model is flexible and can admit additional endogenous variables. In this paper, however, we focus primarily on the estimation of local spillovers.<sup>12</sup>

Note that estimation is hierarchical, in that instrumental variable (IV) is employed in the second stage of the procedure. To avoid confusion in wording, we use the term “instrumental variable (IV) estimation” instead of “two-stage least squares”. Whenever the phrase “two-stage” or “second stage” is used, it refers uniquely to the second stage of the estimation procedure. The procedure of the full estimation is shown in the following structure.

The procedure  $\begin{cases} \text{first stage} & : \text{estimate } \beta_1, \alpha_1, \lambda, \gamma \text{ by a multinomial logit model} \\ \text{second stage} & : \text{estimate } \beta_0, \alpha_0 \text{ by IV } \begin{cases} \text{auxiliary regression} \\ \text{use predicted value as the instrument} \end{cases} \end{cases}$

### 3.3 Instrumental variables

The first instrumental variable is derived from the internal logic of the sorting model. Throughout the analysis, all agents play a static simultaneous-move game. Given the utility function in equation (1), the probability,  $Prob_{mj}$ , that migrant  $m$  chooses alternative  $j$  can be written as a function of all regional and individual characteristics (observed or unobserved). Aggregating these probabilities over all migrants yields the

<sup>12</sup>The variables of local amenities could be possibly endogenous as well if the local government is adjusting the local provision in response to the demand of incoming migrants. It will still lead to overestimated coefficients; however, the local provision of medical and educational resources in China is largely supply-driven, not demand-driven. Even if it is demand-driven, it appeals to the local residents. The endogeneity of amenities might be less of a concern here.

share of individuals choosing province  $j$  as follows:

$$\sigma_j = \frac{1}{M} \sum_{m=1}^M Prob_{mj} = \frac{1}{M} \sum_m \frac{\exp(\mathbf{Z}'_j \beta_m + \alpha_m \sigma_j + \mathbf{P}'_{ij} \gamma + \eta_j + \varepsilon_{mj})}{\sum_{j=1}^J \exp(\mathbf{Z}'_j \beta_m + \alpha_m \sigma_j + \mathbf{P}'_{ij} \gamma + \eta_j + \varepsilon_{mj})}. \quad (6)$$

The sorting equilibrium is the outcome that every location decision is optimal given the location decisions of all others. We need to find a variable that predicts the share of individuals choosing a specific location that is not correlated with the unobserved characteristics of the location. Assuming there are no local spillovers, the artificial instrument  $\sigma^*$  constructed arises naturally out of the following equilibrium condition: by imposing  $\eta = 0$  and  $\sigma = 0$ , we solve for the predicted  $\sigma_j^*$  that would clear the market when only the observed province characteristics are considered by the migrants.

$$\sigma_j^* = \frac{1}{M} \sum_m \frac{\exp(\mathbf{Z}'_j \beta_m + \mathbf{P}'_{ij} \gamma)}{\sum_{j=1}^J \exp(\mathbf{Z}'_j \beta_m + \mathbf{P}'_{ij} \gamma)} \quad \text{for } m \text{ in } j. \quad (7)$$

This approach requires an iterating procedure for both stages. The initial values of  $\beta_0$  are obtained by estimating equation (5) using ordinary least squares (OLS). The estimates of  $\beta_0$ , together with  $\beta_1$  and  $\gamma$  from equation (4), are then used to calculate a new  $\sigma^*$  under equilibrium conditions (7) after imposing  $\eta = 0$  and  $\sigma = 0$  for all  $j$ . The new vector we solve is then used as an IV for  $\sigma_j$  in equation (5).  $\beta_0$  coefficients are then updated as IV estimation coefficients. The new values of  $\beta_0$  are plugged back into the equilibrium conditions (7) in the same way as before, and this process is repeated until the instrument stabilizes.

The second instrumental variable for  $\sigma_j$  is the provincial highway passenger flow in 1979 and relies on the unique economic history of China. Voluntary migration was not allowed before 1978 under the planned economy. After the economic reform, the market gradually became liberalized, as did migration flows. The provincial highway passenger flow in 1979 reflects the initial status of highway connections, which is not likely to be driven by past or current demand for labor in cities. The relevance of the instrument lies in the fact that, when the restriction on mobility was loosened, people more often moved to provinces with a higher level of accessibility (mostly by intercity coaches). The multiplying effect of local spillovers attracts more migrant workers and hence, continues to affect the present attractiveness of provinces.



## 4 Data and Variables

The full estimation of the model requires the following three types of data: individual socio-demographic characteristics,  $X_m$ , origin-destination pair information,  $P_{ij}$ , and province attributes,  $Z_j$ .

Our main dataset is the China Migrants Dynamic Survey (CMDS), which was conducted by the National Health Commission of the People's Republic of China in 2014. Migrants are defined as those who have moved across county boundaries without a local hukou and have been living in their current city for more than one month. The survey selected respondents aged 16 to 59 by using multi-stage, clustered sampling based on the probability proportionate to size technique. One member of the household (usually the household head) was chosen to answer questions for the household. The survey officially started in 2009 and is implemented annually. After two years of piloting the survey, 2011 is the first year with a representative sample. We only use the 2014 wave for the analysis because pooling the surveys from 2011 to 2014 might miscount migrants who moved in later years. In 2014, 200,937 households were interviewed in 32 provincial-level administrative divisions in total.<sup>13</sup>

To analyze the post-GFC migration pattern in China, the CMDS is advantageous compared with many datasets frequently used by other research on migration. In terms of spatial coverage and number of observations, it outperforms other household surveys such as Rural Urban Migration in China (RUMiC) and China Labor Force Dynamics Survey (CLDS).<sup>14</sup> Zhang and Zhao (2013) uses the RUMiC survey in 2008, and analyzes the location choices of 5,000 rural-urban migrant households. Liu et al. (2015) uses the CLDS in 2012 and only 2,386 migrants are identified. The only database that contains more observations than the CMDS is the Census or Inter-Census Survey, as used by Fu and Gabriel (2012) and Poncet (2006). The most up-to-date Inter-Census Survey conducted in 2015 should be an appropriate dataset for this topic. Unfortunately, with regard to the question to migrants, “how long have you left the place of hukou?”, the response options make it impossible to identify migrants who moved between the years 2008 and 2010, a critical period to study the immediate migration responses after the

---

<sup>13</sup>The data access and detailed documents can be found at <http://www.chinaldrk.org.cn/>.

<sup>14</sup>The RUMiC is launched by the Australian National University and is targeted at rural-urban migrant workers in destination cities. The CLDS is launched by Sun Yat-Sen University and it is the first national longitudinal social survey targeted at the labor force in China.

GFC.

The CMDS provides rich information on individual demographic characteristics and asks retrospective questions about the respondent's migratory history. We are interested in rural-urban migrant workers who hold a rural hukou but work in cities. Our sample observations satisfy four criteria. First, the workers migrated between the years 2008 and 2014. This time range aims to put the focus on the post-GFC choice pattern for destinations. Second, the respondents are not accompanying migrants, such as children, spouses, or relatives of a specific family member. Third, they were at least 16 years old at the time of migration. This is to ensure that they are capable of independently making a migration decision on the basis of their own utility. Fourth, they are not self-employed entrepreneurs. These criteria leave us with a sample of 58,595 rural-urban migrant workers. The destinations of the sample cover the following 26 provincial-level administrative divisions: Anhui, Beijing, Chongqing, Fujian, Gansu, Guangdong, Guangxi, Guizhou, Hebei, Heilongjiang, Henan, Hubei, Human, Inner Mongolia, Jiangsu, Jiangxi, Jilin, Liaoning, Shaanxi, Shandong, Shanghai, Shanxi, Sichuan, Tianjin, Xinjiang, Yunnan and Zhejiang.<sup>15</sup>

Table 1 presents the descriptive statistics of these migrant workers' individual characteristics. The average age is approximately 31 years old, 61 percent of the sample are male and almost 63 percent report being married. On average, they have ten years of education, which is slightly above middle school level. The average age at migration is 29, with a 90 percent central range of 18 to 46. Of all migration decisions, 53 percent are inter-province.<sup>16</sup>

[TABLE 1 ABOUT HERE]

The individual characteristics at migration in the  $X_m$  vector used for estimating Equation (4) are as follows: male, age, years of education, number of school-aged children (under age 16), and potential income. Years of education is calculated as  $\min[\text{AgeAtMigration}-6, \text{EduYearsAtSurvey}]$ . As six is the starting age for primary schooling, this calculation takes into account the possibility that some migrant workers had

<sup>15</sup>Our analysis is restricted to destination choices in the Mainland of China at the provincial level. Hainan, Ningxia, Qinghai, Tibet, and Xinjiang are removed because the total floating population size reported by the Census is too small.

<sup>16</sup>In the Appendix, Table A1 shows how the variables of the individual characteristics are asked in the questionnaires, including the detailed definitions and coding.

additional schooling after migration. Potential income is a predicted value. The income presented in Table 1 cannot be used directly for the estimation, as we need each migrant's potential income in each province. We run an origin-province-specific OLS regression of income on individual characteristics and use the predicted value as the potential income.

The origin-destination pair variables,  $P_{ij}$ , are supplemented by China Geographic Information System Data<sup>17</sup>, Population Census 2000, and the Language Atlas of China (Wurm et al., 1987). *Distance*: the geographic distance between the centroid of provinces  $i$  and  $j$ . *Co-origin network*: the number of migrants from origin  $i$  divided by the total population of the destination  $j$ . *Linguistic distance*: the linguistic distance measure between the origin and destination, constructed by counting the shared number of linguistic groups based on the Language Atlas of China.<sup>18</sup> We use the dialect spoken in the capital city of the province to represent the language of the province: the linguistic distance measure is equal to 3 if the two dialects do not share features with any common linguistic group; the measure is equal to 2 if the two dialects only share features at the most aggregated level of the linguistic group; the measure is equal to 1 if the two dialects share features in the first and second linguistic groups; finally, the measure is equal to 0 if the two dialects share features with all three levels of the linguistic group. In Table 2, we show in particular the descriptive statistics for inter-provincial migrants, i.e., those migrants who cross provincial borders. On average, co-origin network percentage in destination province is 5 percent. For some origin provinces with large migrant outflows, the value can be as high as 19 percent.

[TABLE 2 ABOUT HERE]

We collect most of the province-specific attributes  $Z_j$  in 2008 from the National Bureau of Statistics (NBS) of the People's Republic of China.<sup>19</sup> *Average income of rural-urban migrants*: calculated as the average income of rural migrants in the destination province, adjusted by the consumption price index. *Employment in secondary industries*: calculated from the total number of workers in mining, manufacturing, electricity and construction. *Employment in tertiary industries*: calculated from the total number of workers in transportation, information technology sectors, retail and wholesale,

<sup>17</sup>See <http://gis2.harvard.edu/resources/data/china-gis-data/> for more details.

<sup>18</sup>See Liu et al. (2015) for further details on this methodology.

<sup>19</sup>Data are publicly available at <http://www.stats.gov.cn/>.

hotels and restaurants, finance, rental and commercial services, research and technical services, public facility management, residential services, educational activities, health and social activities, recreational activities and public administration.<sup>20</sup> *Employment in healthcare*: the number of health staff owned per 10,000 people, which proxies for medical services. *Private education spending*: private funds in education spent per 10,000 people.<sup>21</sup> *Share of migrants among all respondents*  $\sigma_j$ : the share of migrants who choose a specific province among all respondents in the survey, which proxies for local spillovers. The descriptive statistics for the province attributes are shown in Table 3. In particular, the average percentage of migrants choosing a destination is approximately 4 percent. Popular destination provinces can attract as many as 11 percent of all respondents.

[TABLE 3 ABOUT HERE]

## 5 Estimation Results

This section presents and discusses the results from estimating the sorting model for the 26 provinces in our sample from the 2014 CMDS. The model is implemented in two stages. In the first stage, individual characteristics of migrants are interacted with province characteristics to estimate heterogeneous preferences controlling for unobserved province heterogeneity. In the second stage, we can then estimate the relative contributions of province characteristics and local spillovers in determining a location's attractiveness to rural-urban migrants. Lastly, we compare the magnitudes of all attributes using the relative risk ratios and discuss some variables of particular interest. All province characteristics (except for migrant shares  $\sigma$ ) and origin-destination pair variables are standardized.

<sup>20</sup>In addition to employment in levels, we also use employment shares. The results remain consistent.

<sup>21</sup>We use private instead of public funds for education to proxy for local educational resources, because privately run schools are the primary venue for the education of migrant children.

## 5.1 First stage

The upper panel of Table 4 reports estimates for the 36 coefficients ( $\beta_1, \alpha_1$ ) on terms that interact individual characteristics and province characteristics. The lower panel reports three coefficients ( $\gamma$ ) for the origin-destination pair variables from equation (4).

[TABLE 4 ABOUT HERE]

If we read the coefficients by row, the upper panel shows, given a fixed level of a certain province attribute, how much more likely the odds ratio of a province being chosen over another will change in response to different individual characteristics. The first row concerns the local average income of rural-urban migrants. Migrants who migrate at an older age (0.091 for the squared term), who have at least one child of school age (0.06) and who have higher potential earnings are likely to move to provinces with higher income levels on average. The second and the third rows concern employment in secondary and tertiary industries. The estimates in the two rows give almost opposite values, indicating a clear sorting pattern of migrants of different types into the two broad industry categories. The last row is of particular interest to us, as it implies the type of migrants preferring to locate in a province with many other rural migrants. The sorting pattern across provinces with different shares of rural migrants is strong. Local spillovers are significant among males (7.336), older (3.884 for the squared term), less-educated (-0.655), and lower-income (-3.593) workers. In other words, for provinces with an existing share of rural migrant workers, there is a much higher tendency for these groups to move there.

If we read the coefficients by column, it shows that given a fixed value of certain individual characteristics, how much more likely the odds ratio of a province being chosen over another will change in response to different province attributes, *ceteris paribus*. Take columns (2) and (3) for example, workers who migrate at a younger age derive higher utility from opportunities in tertiary industries. In column (4), lower-educated migrant workers derive higher utility from opportunities in secondary industries (-0.113), while the higher-educated derive higher utility from employment in tertiary industries (0.121). In the last column, only potentially high-income earners derive

higher utility from local medical services (0.246) and better educational resources (0.188).

The estimates of  $\gamma$  for the origin-destination pair variables are shown in the lower panel of Table 4. Geographic distance reduces the probability of a potential migrant worker choosing a destination (-0.992). The network share from a migrant's province of origin makes it more attractive for migrant workers to move in (0.445). We see this variable as capturing the effect of social networks formed by migrants from the same place ("Lao Xiang"). Furthermore, larger linguistic distance seems to deter migrant workers from choosing a destination province (-0.19).

Table 5 reports the mean indirect utility,  $\lambda$ , of each province estimated in the first stage as alternative-specific constants. Beijing, Fujian, Guangdong, Jiangsu, Shanghai, and Zhejiang rank the highest and are the most attractive destinations for migrant workers. They are all located in the eastern region. Henan, Sichuan, Anhui, Guangxi, and Shanxi rank lowest by mean indirect utility. The ranking of the provinces is comparable to the ranking of top 20 cities in Xing and Zhang (2017). Except for the two cities of Shenyang and Dalian, all other cities are included in our top six provinces.

[TABLE 5 ABOUT HERE]

## 5.2 Second stage

In Table 6, mean indirect utility,  $\lambda$ , is decomposed into each province's characteristics,  $\beta_0$ , and local spillovers,  $\alpha_0$ , from equation (5).

[TABLE 6 ABOUT HERE]

In column (1), we report OLS estimates over the average income of rural-urban migrants, employment in secondary industries, employment in tertiary industries, employment in healthcare, private education spending and migrant shares  $\sigma$ . A higher local income level for rural migrants, a large employment in secondary industries, better medical services, and a larger population of rural migrants all increase province's

attractiveness. Columns (2) reports the IV estimation using the artificial instrumental variable:  $\sigma^*$  is the equilibrium  $\sigma$  value, which is, by definition, orthogonal to province's attractiveness but relevant to the observed migrant share  $\sigma$ . In column (3), we use the provincial highway passenger flow in 1979 as another instrument for  $\sigma$ . As expected, a higher volume of highway passenger inflow is positively correlated with  $\sigma$  due to the convenience of transportation. The F-test value in the first-stage regression is 10.87, implying a borderline weak instrument. However, both sets of IV estimation results are consistent with those of the OLS estimation.

The effects of local spillovers are shown in the last row of Table 6. As the value of the migrant share is used to calculate the artificial instrumental variable, we do not standardize this variable. Thus, the magnitude of the coefficient here should be interpreted with caution. Take column (3), for example, one standard deviation of the migrant share variable (0.02) leads to a value of 1.466 ( $\approx 0.02 \times 73.291$ ) for the mean indirect utility  $\lambda$ , which is as strong as the impact of employment in the secondary industries on  $\lambda$ . The more migrants that move to a specific province, the more likely it is that a potential migrant worker will choose that location.

The estimate for educational resources is not statistically significant throughout the specifications. This looks unintuitive at first sight. However, it should be noted that the measure for educational resources is private funds instead of the government financial funding for education. As the children of rural migrants are denied access to public schools in cities, they can only go to low-cost private schools, which are of inferior quality. Our results imply that local educational funding does not add to province's attractiveness at all in the second stage. Returning to the results for the first stage, the interaction term between the number of school aged children and the private funding of education is not significant either.

### 5.3 Comparison by relative risk ratios

In the simplest version of a multinomial logit model with only one predictor,  $z$ , for many alternatives, we call the probability of province  $j$  being chosen relative to the probability of baseline province  $k$  being chosen as the relative risk (commonly known as odds ratio). When certain predictor of interest,  $z$ , changes to  $z + 1$ , ceteris paribus,



the relative risk ratio (RRR) is calculated as follows:

$$RRR = \frac{Pr(j|z+1)/Pr(k|z+1)}{Pr(j|z)/Pr(k|z)} = \exp(b), \quad (8)$$

where  $b$  is the coefficient for predictor  $z$ . One unit increase in  $z$  leads to a composite impact on the relative risk as shown above.  $b > 0$  is equivalent to  $RRR > 1$ , which implies that the probability of  $j$  being chosen over  $k$  is getting larger due to the 1 unit increase in predictor  $z$ . In this way, the main results in all tables can be easily summarized with  $\exp(b)$ .

The pull effect of jobs is mainly due to average income and employment in secondary industries. In column (3) of Table 6, if we increase province  $j$ 's employment in secondary industries by one standard deviation (127 million positions), the RRR is  $3.1 \approx \exp(1.145)$  times greater, ceteris paribus, i.e., the odds of province  $j$  being chosen versus the baseline province being chosen are 3.1 times greater. Equivalently, if a province has a higher income level by one standard deviation (366 RMB), the RRR is  $2.2 \approx \exp(0.774)$  times greater. Better medical resources (22 more healthcare staff per 10,000 people) turn out to be attractive for rural migrants (RRR=2.9), but we do not find such a pull effect for educational resources.

The overall share of migrant workers increases the province's attractiveness with a similar magnitude. Here, we are mostly interested in how the group of rural migrants in general differ from the group of migrants coming from the same province in affecting one's destination choice. RRR is  $1.6 \approx \exp(0.445)$  times greater with one standard deviation increase (5 percent) in the population share of origin province at the destination province, whereas RRR is  $4.3 \approx \exp(0.02*73.291)$  times greater with one standard deviation increase (2 percent) in the share of migrants moving to a destination among all respondents.

## 6 Robustness Checks

We perform several robustness checks to validate our main results. First, we check whether the preference for amenities is confounded by possible job opportunities in medical- and education- related jobs. Second, we extend our model by considering

the possible presence of spatial autocorrelation. Third, we replace employment in secondary industries with employment in manufacturing only. Fourth, we implement the analysis at a finer regional level, i.e., cities. Fifth, we divide the sample into a more-educated group and a less-educated group and check how preferences differ.

[TABLE 7 ABOUT HERE]

First, we remove from the sample respondents who work in medical- and education-related jobs. Provinces with better public services are usually associated with a high demand for workers in these jobs. The estimate for medical services might imply a combined effect of amenity sorting and job sorting. Column (1) of Table 7 shows that results are consistent with previous findings, implying that the effect of these province variables is not confounded by other channels. The demand for public services and amenities plays a major role here.

Second, we use a spatial autoregressive model (Anselin, 1988) to deal with the possible presence of spatial autocorrelation, namely, the impact of these province attributes can possibly extend over geographic boundaries. Thus, a province could generally be regarded as more attractive when it is closer to an attractive province. The specification is written as follows:

$$\lambda = \rho \mathbf{W} \lambda + \beta_0 \mathbf{Z} + \alpha_0 \sigma + \eta, \quad (9)$$

where  $\mathbf{W}$  is a spatial matrix measuring the proximity of neighboring provinces. Two common types of such matrices are (1) the inverse distance matrix in which the weight is calculated by the inverse of the geographic distance between provinces and (2) the contiguity matrix in which the weight simply indicates whether spatial units share a geographic boundary. Columns (2) and (3) of Table 7 show the contribution when the attractiveness of neighboring provinces is controlled for. Whether using the inverse distance matrix or the contiguity matrix, spatial autocorrelation seems trivial at the provincial level.

Third, the provinces receiving the most rural migrant workers are also more economically developed and focus on labor-intensive industries, especially the manufacturing

industry. We look particularly at how the employment in manufacturing only affects the migrants' spatial sorting. Column (4) of Table 7 shows that the estimate for the employment in manufacturing becomes slightly smaller, while the other estimates remain relatively similar in magnitude. The contribution of manufacturing alone makes up approximately 2/3 of the contribution from secondary industries.

Fourth, we use a more disaggregated geographical unit and test for “sorting by city” instead of “sorting by province”. Although we have considered the differences between provincial units, there could be considerable heterogeneity within provinces. To address this issue, we restrict the sample to the rural migrant workers originally from Sichuan province, which is one of the largest rural migrant sending provinces. Among the top 40 cities chosen as destinations, we implement our analysis. Eighty percent of migrants from Sichuan migrated to these 40 cities, ensuring the validity of the choice set. We also adjust the number of cities to 30 or 50, and the results are similar.<sup>22</sup> Column (5) of Table 7 shows that the signs of these estimates remain the same as the main results. The average wage level (0.238) and the employment in secondary industries (0.296) contribute to city's attractiveness. Medical resources at the city level (0.386) are much more important than at the provincial level. The impact of the migrant share at the city level on the mean indirect utility is 0.237 ( $\approx 0.03 \times 7.895$ ).<sup>23</sup> The effect of local spillovers is twice as large as the impact of employment in the secondary industries, implying that the effect of local spillovers is much stronger with a more disaggregated regional level.

Lastly, we divide the sample into a more-educated group (years of schooling  $> 9$ ) and a less-educated group (years of schooling  $\leq 9$ ) to further investigate the role of heterogeneous preferences. Nine years of schooling is a critical point of having finished compulsory education in China, and hence, there might be systemic differences between the two groups. The more-educated group only accounts for one third of the sample. They appear to be younger and show a higher willingness to settle in the destination in the future. Table 8 summarizes some key estimates of the subsamples, both in the first and second stages. The heterogeneous impacts in the first stage are rather similar between the two groups, except that local spillovers are especially strong for

---

<sup>22</sup>Our proxy for the educational resources at the city level is slightly different from the variables at the provincial level due to data limitations. The educational resources are now measured by the total number of students enrolled in primary schools.

<sup>23</sup>The migrant share is not standardized, so the impact should be calculated by multiplying one standard deviation of the variable and the coefficient.

the less-educated group. In the second stage, the estimate for the less-educated group remains similar to the main results. However, the first column for the more-educated group is the only context where the educational resources turn out to be important for province's attractiveness. Furthermore, employment in secondary industries fails to be an attractor for this particular group, perhaps because they can opt to become self-employed instead. At the province level, the impact of migrant share on the more-educated group is lower than that on the less-educated group, which substantiates our main results.

[TABLE 8 ABOUT HERE]

## 7 Conclusion

Since the reform of China's economy in 1978, coastal provinces in the east have traditionally received large migrant inflows drawn from rural areas in the interior of the country. We identify how following the global financial crisis, rural-urban migration in China is increasingly an intra-province phenomenon, especially for provinces in the middle and western regions. In this paper, we examine what might be causing these changes in the spatial distribution of migration flows by analyzing the sorting behavior between the years 2008 and 2014 of 58,595 migrant workers in a large-scale government survey. Three sets of factors are studied: jobs, amenities, and local spillovers. We employ the equilibrium-sorting model developed by Bayer and Timmins (2007). This methodology follows the conventional discrete choice framework, but allows us to estimate spillover effects using location decisions when local attributes are unobserved. Using this approach, similar to previous studies, we find that rural-urban migrants prefer destinations close to their province of origin, locations in which many people from their own province reside, and destinations where the dialect is similar.

New in this paper, after controlling for the aforementioned factors, we find a separate and strong preference for co-locating with a large population of rural migrants, regardless of their origins. What these migrants all have in common is that they lack the urban hukou that would give them access to the same education and employment opportunities as the local residents. The effect is particularly strong among older, less-educated, and lower-income rural-urban migrant workers. We take this as evidence in support

of a channel where a network of rural migrants might help to reduce the effects of systemic discrimination on these more vulnerable groups. We cannot test labor market pooling and housing competition channels directly due to missing information about job mobility and housing quality. What we can claim is that the social and economic benefits brought by a large community of rural migrants appear to be much higher than the potential costs in various aspects. The results remain robust when we take into account labor supply-driven migration, spatial autocorrelation between provinces, different industry definitions, and regional differences within provinces.

Our study can inform urbanization policies in China and the future distribution of rural-urban migrants.

First, for the foreseeable future, the size of the rural-migrant community will still be an important factor in destination choices due to institutional barriers. Although there have been several small breakthroughs in hukou reform since the 1980's due to a decentralized process of hukou management, rural-urban migrant workers have not benefited much from these changes. Some provinces have removed the distinction between rural and urban hukou within their own localities (Song, 2014). These changes mostly benefit people who already have local hukou in those places. As the current hukou system places barriers on rural-urban migrants based on two classifications, i.e., the hukou type (rural vs. urban) and the hukou location (local vs. non-local), the efforts to abolish the distinction between rural and urban hukou do not equate with efforts to address the distinction between local and non-local hukou. Migrant workers are still not entitled to the same benefits as local workers, and there is a long way to go.

In addition, our results suggest that ongoing industrial upgrading and transfer policies in China might exert a sizable effect in redirecting rural-urban migration flows westward in the future. Following the global financial crisis, eastern provinces have developed more capital-intensive sectors and manufacturing firms have moved labor-intensive plants to the middle and western regions (Ang, 2018; Meng, 2014; Peng, 2015). As much rural-urban migration originates from provinces in these regions, we expect intra-province migration to grow stronger as migrants are more able to find jobs in the nearby cities. This westward movement will be accelerated as spillover effects draw in still more migrants, in particular from older, less-educated, and lower-income demographic groups.

Lastly, our results provide guidance to local policy makers that the lack of schooling access for rural migrants is the most urgent problem to tackle. Migrant destination choice is almost inelastic to changes in local private educational funding, which currently provides the primary access to education for migrant children. We find that only more-educated migrant workers are able to consider educational resources in location choices. The lack of schooling access for rural migrants will potentially be a crucial determinant for whether migration flows to cities will be temporary or permanent in nature.

This study can be extended in many directions in the future. First, we study how migrants choose destinations, and the results are relevant for people who have decided to leave rural areas and work in cities. The limitation of discrete choice models in tackling the selection problem of the target group can be weakened to a certain extent if data about the migrants' pre-migration characteristics and counterpart groups in rural areas are available. Second, the current estimates only roughly consider the impact of employment in secondary industries, and the variable is not merely determined by the demand-factors of labor. An analysis of exogenous industrial policies at a more disaggregated level is important for policy-making. For example, the change in the content of new vacancies due to industrial upgrading could help researchers understand the sorting patterns of migrant workers to a greater extent. Third, as most of the industrial transfer policies go against economies of scale, the economic efficiency of these policies needs to be evaluated apart from the impact on migration flows. Due to limitations in the data currently available, these details are not discussed in this paper.

## References

- Akay, A., O. Bargain, and K. F. Zimmermann (2012). 'Relative concerns of rural-to-urban migrants in China'. *Journal of Economic Behavior & Organization* 81, pp. 421–441.
- Ang, Y. Y. (2018). 'Domestic flying geese: industrial transfer and delayed policy diffusion in China'. *China Quarterly* 234.October 2015, pp. 420–443.
- Anselin, L. (1988). *Spatial Econometrics: Methods and Models*. Dordrecht: Kluwer Academic.

- Åslund, O. (2005). 'Now and forever? Initial and subsequent location choices of immigrants'. *Regional Science and Urban Economics* 35, pp. 141–165.
- Bayer, P. and C. Timmins (2005). 'On the equilibrium properties of locational sorting models'. *Journal of Urban Economics* 57, pp. 462–477.
- (2007). 'Estimating equilibrium models of sorting across locations'. *The Economic Journal* 117.March, pp. 353–374.
- Bayer, P., R. Mcmillan, and K. Rueben (2004). 'An equilibrium model of sorting in an urban housing market'. *NBER No.* 10865.
- Berry, S., J. Levinsohn, and A. Pakes (1995). 'Automobile prices in market equilibrium'. *Econometrica* 63.4, pp. 841–890.
- Buckley, F. (1996). 'The political economy of immigration policies'. *International Review of Law and Economics* 16.1, pp. 81–99.
- Chan, K. W. (2010). 'The global financial crisis and migrant workers in China: 'there is no future as a labourer; returning to the village has no meaning''. *International Journal of Urban and Regional Research* 34.3, pp. 659–677.
- (2012). 'Migration and development in China: trends, geography and current issues'. *Migration and Development* 1.2, pp. 187–205.
- Chan, K. W. and W. Buckingham (2008). 'Is China abolishing the hukou system?' *China Quarterly* 195.Sep, pp. 582–606.
- Chen, Y. (2005). 'Peasant-labor: system and identity'. *Sociological Studies (in Chinese)* 3, pp. 119–132.
- Chen, Y. and S. Feng (2017). 'Quality of migrant schools in China: evidence from a longitudinal study in Shanghai'. *Journal of Population Economics* 30.3, pp. 1007–1034.
- Combes, P.-P., S. Démurger, and S. Li (2015). 'Migration externalities in Chinese cities'. *European Economic Review* 76, pp. 152–167.
- Démurger, S., M. Gurgand, S. Li, and X. Yue (2009). 'Migrants as second-class workers in urban China? A decomposition analysis'. *Journal of Comparative Economics* 37.4, pp. 610–628.
- Fu, Y. and S. A. Gabriel (2012). 'Labor migration, human capital agglomeration and regional development in China'. *Regional Science and Urban Economics* 42.3, pp. 473–484.
- Giles, J., X. Meng, S. Xue, and G. Zhao (2018). 'Does information influence the social insurance participation decision of China's rural migrants?' *Working paper*.
- Hu, F., Z. Xu, and Y. Chen (2011). 'Circular migration, or permanent stay? Evidence from China's rural-urban migration'. *China Economic Review* 22.1, pp. 64–74.



- Kong, S. T., X. Meng, and D. Zhang (2010). 'The global financial crisis and rural-urban migration'. *China Update*. Ed. by R. Garnaut, L. Song, and W. T. Woo. Brookings Institution Press, pp. 241–265.
- Kuang, L. and L. Liu (2012). 'Discrimination against rural-to-urban migrants: the role of the Hukou system in China'. *PLoS ONE* 7.11.
- Lai, F., C. Liu, R. Luo, L. Zhang, X. Ma, Y. Bai, B. Sharbono, and S. Rozelle (2014). 'The education of China's migrant children: the missing link in China's education system'. *International Journal of Educational Development* 37, pp. 68–77.
- Levkovich, O. and J. Rouwendal (2014). 'Location choices of highly educated foreign workers: the importance of urban amenities'. *Tinbergen Institute Discussion Paper* 093/VIII.
- Liang, Z. and M. J. White (1997). 'Market Transition, Government Policies, and Interprovincial Migration in China: 1983-1988'. *Economic Development and Cultural Change* 45.2, pp. 321–339.
- Liang, Z., Z. Li, and H. Morooka (2017). 'Differential effects of migration networks: The case of China'. *Chinese Journal of Sociology* 3.3, pp. 354–378.
- Lin, J. Y., G. Wang, and Y. Zhao (2004). 'Regional inequality and labor transfers in China'. *Economic Development and Cultural Change* 52.3, pp. 587–603.
- Liu, Y. and W. Xu (2017). 'Destination choices of permanent and temporary migrants in China, 1985-2005'. *Population, Space and Place* 23.1.
- Liu, Y., X. Xu, and Z. Xiao (2015). 'The pattern of labor's cross-dialects migration'. *Economic Research Journal (in Chinese)* 10, pp. 134–162.
- Lu, J. and Z. Tao (2009). 'Trends and determinants of China's industrial agglomeration'. *Journal of Urban Economics* 65.2, pp. 167–180.
- Meng, X. (2014). 'China's labour market tensions and future urbanisation challenges'. *Deepening Reform for China's Long-term Growth and Development*. Ed. by L. Song, R. Garnaut, and F. Cai. ANU Press.
- Meng, X. and J. Zhang (2001). 'The two-tier labor market in urban China-occupational segregation and wage differentials between urban residents and rural migrants in Shanghai'. *Journal of Comparative Economics* 29.3, pp. 485–504.
- Meng, X., K. Shen, and S. Xue (2013). 'Economic reform, education expansion, and earnings inequality for urban males in China, 1988-2009'. *Journal of Comparative Economics* 41.1, pp. 227–244.
- Peng, G. (2015). 'The matching of skills to tasks, labor migration and Chinese regional income disparity'. *Economic Research Journal (in Chinese)* 1, pp. 99–110.
- Poncet, S. (2006). 'Provincial migration dynamics in China: borders, costs and economic motivations'. *Regional Science and Urban Economics* 36, pp. 385–398.

- Rozelle, S., L. Guo, M. Shen, A. Hughart, and J. Giles (1999). 'Leaving China's farms: survey results of new paths and remaining hurdles to rural migration'. *The China Quarterly* 158, Jun, pp. 367–393.
- Song, Y., Y. Zenou, and C. Ding (2008). 'Let's not throw the baby out with the bath water: the role of urban villages in housing rural migrants in China'. *Urban Studies* 45.2, pp. 313–330.
- Song, Y. (2014). 'What should economists know about the current Chinese hukou system?' *China Economic Review* 29, pp. 200–212.
- Su, Y., P. Tesfazion, and Z. Zhao (2018). 'Where are the migrants from? Inter- vs. intra-provincial rural-urban migration in China'. *China Economic Review* 47, pp. 142–155.
- Tiebout, C (1956). 'A pure theory of local expenditures'. *The Journal of Political Economy* 64.5, pp. 416–424.
- Van Duijn, M. and J. Rouwendal (2013). 'Cultural heritage and the location choice of Dutch households in a residential sorting model'. *Journal of Economic Geography* 13.3, pp. 473–500.
- Wang, F. and X. Zuo (1999). 'Inside China's cities: institutional barriers and opportunities for urban migrants'. *The American Economic Review* 89.2, pp. 276–280.
- Wang, Z., T. De Graaff, and P. Nijkamp (2016). 'Cultural diversity and cultural distance as choice determinants of migration destination'. *Spatial Economic Analysis* 11.2, pp. 176–200.
- (2018). 'Barriers of culture, networks, and language in international migration : a review'. *REGION* 5.1, pp. 73–89.
- Wu, W. (2004). 'Sources of migrant housing disadvantage in urban China'. *Environment and Planning A* 36, pp. 1285–1304.
- Wurm, S. A., R. Li, T. Baumann, and M. W. Lee (1987). *Language Atlas of China*. Hong Kong: Longman.
- Xia, Y. and M. Lu (2015). 'Move between cities: empirical evidence on the effect of public service on labour migration'. *Management World (in Chinese)* 10, pp. 78–90.
- Xing, C. and J. Zhang (2017). 'The preference for larger cities in China: evidence from rural-urban migrants'. *China Economic Review* 43, pp. 72–90.
- Zavodny, M. (1999). 'Determinants of recent immigrants' locational choices'. *International Migration Review* 33.4, pp. 1014–1030.
- Zhang, J. and Z. Zhao (2013). 'Measuring the income-distance tradeoff for rural-urban migrants in China'. *IZA Discussion Paper* 7160.
- Zhao, Y. (1999). 'Leaving the countryside: rural-to-urban migration decisions in China'. *The American Economic Review* 89.2, pp. 281–286.

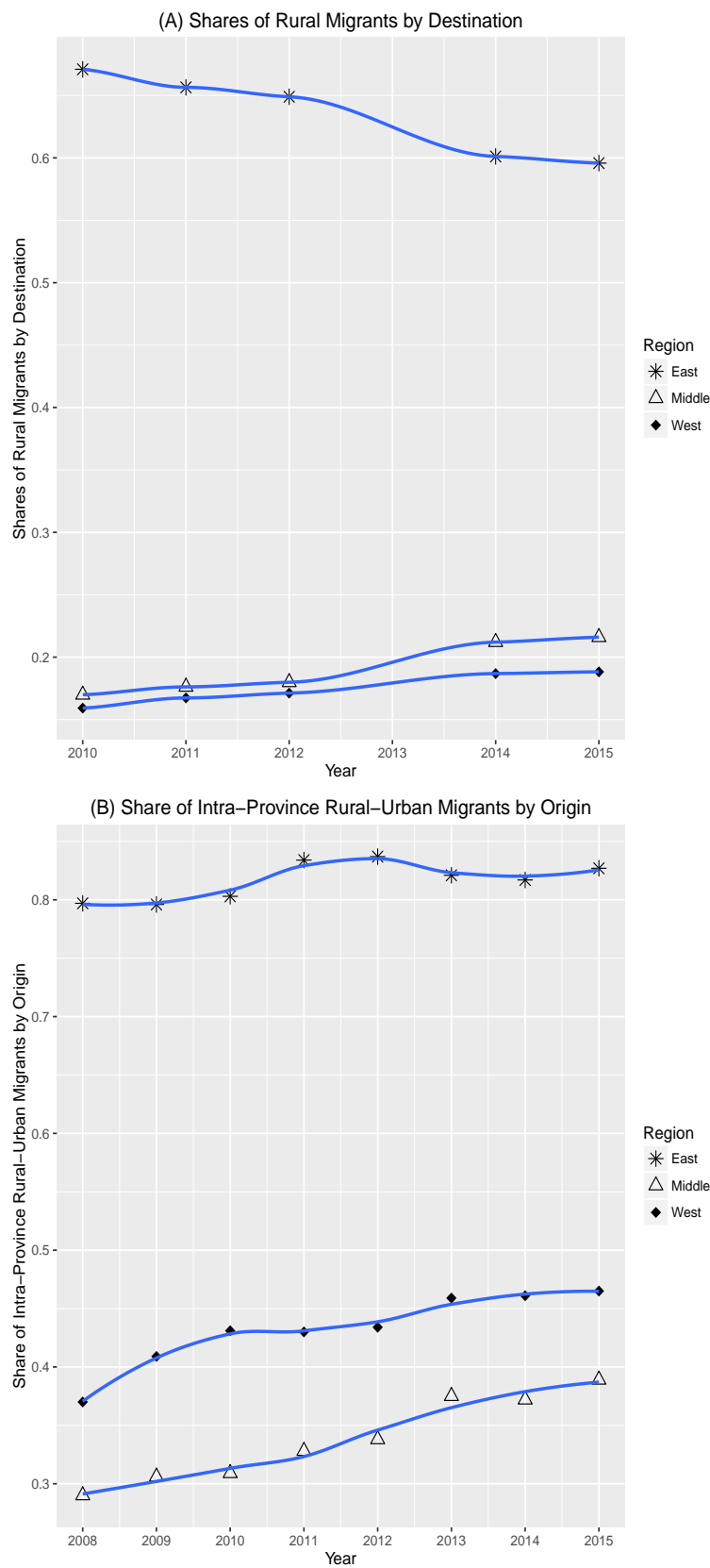
Accepted Article

Zhao, Y. (2003). 'The role of migrant networks in labor migration: the case of China'. *Contemporary Economic Policy* 21.4, pp. 500–511.

Zhao, Z. (2005). 'Migration, labor market flexibility, and wage determination in China: a review'. *The Developing Economies* June, pp. 285–312.

Zheng, S., F. Long, C. C. Fan, and Y. Gu (2009). 'Urban villages in China: a 2008 survey of migrant settlements in Beijing'. *Eurasian Geography and Economics* 50.4, pp. 425–446.

Figure 1: The Dynamics of Rural Migration after the Global Financial Crisis



Note: The share of rural migrants is not reported for 2008, 2009 and 2013 in the annual reports.  
Source: Rural Workers Monitoring Survey Reports 2008-2015.

## Tables

Table 1: Descriptive Statistics for Rural-Urban Migrant Workers

Statistic	Mean	St. Dev.	Min	Max
Current age	31.393	9.090	16	60
Male	0.612	0.487	0	1
Married	0.628	0.483	0	1
Being minority	0.062	0.241	0	1
Current educated years	9.957	2.674	0	16
Number of all children	0.845	0.840	0	5
Number of school-aged children	0.669	0.768	0	5
Work in secondary industries	0.476	0.499	0	1
Work in tertiary industries	0.524	0.499	0	1
Income in RMB	3,132.729	1,397.524	0	24,000
Years since migration	2.098	1.770	0	6
Age at migration	29.295	8.914	16	60
Educated years at migration	9.922	2.634	0	16
Inter-provincial migration	0.527	0.499	0	1

Source: CMDS 2014; sample size 58,595.

Table 2: Descriptive Statistics for Origin-Destination Province Pairs for Rural-Urban Migrants

Statistic	Mean	St. Dev.	Min	Max
Geographic distance (km)	793.04	500.03	115.46	3,653.41
Co-origin network	0.05	0.05	0.00	0.19
Linguistic distance	2.34	0.93	0	3

Note: This table excludes intra-province migrants. Co-origin network is the number of migrants from the origin province,  $i$ , as a share of population in destination province,  $j$ .

Source: China Geographic Information System Data; Population Census 2000; the Language Atlas of China.

Sample size: 30,875

Table 3: Descriptive Statistics for Provinces

Statistic	Mean	St. Dev.	Min	Max
Average income of rural-urban migrants (RMB)	3,282	366	2,637	4,015
Employment in secondary industries ( $\div 10,000$ )	201	127	68	498
Employment in tertiary industries ( $\div 10,000$ )	238	105	106	500
Employment in healthcare (per 10,000 people)	43	22	22	122
Private education funds (per 10,000 people in RMB)	50,727	33,061	4,880	140,470
Share of migrants among all respondents	0.038	0.023	0.016	0.112

Note: 26 provinces are represented in the sample.

Source: NBS 2008 and CMDS 2014.

Table 4: First-Stage Estimation Results

	Male	AAM	AAM <sup>2</sup>	EduYears At Migration	School-aged Children	Potential Income
$\beta_1$						
Average income	-0.186*** (0.031)	-0.070*** (0.008)	0.091*** (0.011)	-0.002 (0.004)	0.060*** (0.012)	0.199*** (0.035)
Employment in secondary industries	-0.530*** (0.043)	-0.020 (0.015)	0.006 (0.021)	-0.113*** (0.007)	0.005 (0.022)	0.478*** (0.044)
Employment in tertiary industries	0.412*** (0.034)	0.073*** (0.011)	-0.090*** (0.015)	0.121*** (0.005)	-0.066*** (0.014)	-0.596*** (0.043)
Employment in healthcare	-0.216*** (0.028)	-0.018** (0.008)	0.019 (0.012)	-0.058*** (0.004)	-0.003 (0.011)	0.246*** (0.032)
Private education spending	-0.155*** (0.022)	-0.018*** (0.006)	0.021** (0.009)	-0.004 (0.003)	0.011 (0.008)	0.188*** (0.027)
$\alpha_1$						
Migr. Share $\sigma$	7.336*** (0.927)	-2.683*** (0.385)	3.884*** (0.573)	-0.655*** (0.169)	3.087*** (0.632)	-3.593*** (0.823)
$\gamma$						
Geographic Distance	-0.992*** (0.012)	Co-origin Network	0.445*** (0.005)	Linguistic Distance	-0.190*** (0.008)	
Individuals	58,595	Provinces	26			

Note: \*  $p < 0.1$ ; \*\*  $p < 0.05$ ; \*\*\*  $p < 0.01$ .

Standard errors are in the parentheses.

Individual variables are listed in the top: male, age at migration, age at migration squared divided by 100, education years at migration, number of children at school age, and potential income divided by 1000.

Province variables are listed in the first column: average income of rural-urban migrants (RMB), employment in secondary industries ( $\div 10,000$ ), employment in tertiary industries ( $\div 10,000$ ), employment in healthcare (per 10,000 people), private education spending (per 10,000 people in RMB), and share of migrants among all respondents.

The origin-destination pair variables are listed at the bottom of the table: geographic distance, percentage of migrants from one's origin province, and linguistic distance.



Table 5: First-Stage Estimation Results, Continued

Province	$\lambda$	Province	$\lambda$
Anhui	0.00	Inner Mongolia	1.81
Beijing	4.41	Jiangsu	6.80
Chongqing	1.20	Jiangxi	2.14
Fujian	5.17	Jilin	0.40
Gansu	1.84	Liaoning	2.59
Guangdong	5.87	Shaanxi	2.30
Guangxi	0.05	Shandong	2.00
Guizhou	0.06	Shanghai	5.47
Hebei	0.84	Shanxi	0.16
Heilongjiang	0.22	Sichuan	-0.48
Henan	-0.39	Tianjin	4.11
Hubei	1.10	Yunnan	0.33
Hunan	0.70	Zhejiang	9.60

Note:  $\lambda$  is the mean indirect utility for each province.  
Anhui is the reference location with mean indirect utility 0.

Table 6: Second-Stage Estimation Results

	Mean indirect utility $\lambda$		
	<i>OLS</i>	<i>instrumental variable</i>	
	(1)	(2)	(3)
Constant	−0.831** (0.333)	−1.451*** (0.396)	−0.575 (0.622)
$\beta_0$			
Average income	0.767*** (0.118)	0.746*** (0.129)	0.774*** (0.126)
Employment in secondary industries	0.987*** (0.265)	0.603* (0.306)	1.145** (0.426)
Employment in tertiary industries	−1.456*** (0.197)	−1.320*** (0.219)	−1.513*** (0.233)
Employment in healthcare	1.028*** (0.137)	0.940*** (0.152)	1.063*** (0.162)
Private education spending	0.058 (0.110)	0.045 (0.121)	0.063 (0.115)
$\alpha_0$			
Migr. Share $\sigma$	79.896*** (8.230)	96.003*** (9.889)	73.291*** (16.073)
Artificial instrument	-	Yes	No
Passenger instrument	-	No	Yes
F-test of exclusion	-	38.04	10.87
Observations	26	26	25
R <sup>2</sup>	0.969	0.963	0.968
Adjusted R <sup>2</sup>	0.959	0.951	0.957
Residual Std. Error	0.520 (df = 19)	0.570 (df = 19)	0.543 (df = 18)

Note: \*p<0.1;\*\*p<0.05;\*\*\*p<0.01

Table 7: Robustness Checks

	Mean Indirect Utility				
	Excl. labor supply	Spatial	Spatial	Manu. only	City reg.
	(1)	(2)	(3)	(4)	(5)
Average income	0.739*** (0.134)	0.769*** (0.122)	0.761*** (0.124)	0.694*** (0.116)	0.238 ** (0.086)
Employment in secondary industries	0.693** (0.318)	0.992*** (0.277)	0.973*** (0.281)	0.667** (0.260)	0.296*** (0.064)
Employment in tertiary industries	-1.385*** (0.227)	-1.460*** (0.206)	-1.436*** (0.226)	-1.403*** (0.163)	-0.300*** (0.045)
Employment in healthcare	0.996*** (0.157)	1.032*** (0.147)	1.014*** (0.157)	0.717*** (0.124)	0.386*** (0.053)
Private education spending	0.075 (0.125)	0.061 (0.118)	0.053 (0.116)	-0.002 (0.107)	-0.099 (0.062)
Migr. Share $\sigma$	95.808*** (10.310)	79.770*** (8.537)	79.570*** (8.611)	97.674*** (9.344)	7.895*** (2.316)
$\rho$ : distance		-0.493 (4.688)			
$\rho$ : contiguity			0.016 (0.080)		
Constant	-1.512*** (0.413)	-0.792 (0.504)	-0.852** (0.358)	-1.623*** (0.373)	-1.193*** (0.090)
Observations	26	26	26	26	40
R <sup>2</sup>	0.962	0.969	0.969	0.971	0.847
Adjusted R <sup>2</sup>	0.950	0.957	0.957	0.962	0.819
Residual Std. Error	0.590	0.534	0.534	0.504	0.387
F Statistic		81.006***	81.130***		30.393***

Note: \*p<0.1,\*\*p<0.05;\*\*\*p<0.01

Table 8: More-Educated vs Less-Educated

Selected Estimates	EduYearsMigr > 9	EduYearsMigr ≤ 9
<b>First Stage</b>		
Male × Migr.Share $\sigma$	10.239*** (1.711)	8.271*** (1.172)
AAM × Migr.Share $\sigma$	-1.559* (0.853)	-1.647*** (0.464)
AAM <sup>2</sup> × Migr.Share $\sigma$	2.705** (1.337)	2.406*** (0.68)
EduYears × Migr.Share $\sigma$	-0.700 (0.486)	-1.238*** (0.396)
School-aged children × Migr.Share $\sigma$	3.966*** (1.325)	3.050*** (0.765)
Potential income × Migr.Share $\sigma$	-8.726*** (1.417)	-7.506*** (1.01)
<b>Second Stage</b>		
Average income	0.64*** (0.13)	0.90*** (0.17)
Employment in secondary industries	-0.33 (0.30)	0.70* (0.40)
Employment in tertiary industries	-0.90*** (0.22)	-1.36*** (0.28)
Employment in healthcare	0.01 (0.15)	0.36* (0.20)
Private education spending	0.28** (0.12)	-0.11 (0.15)
Migr. Share $\sigma$	97.69*** (9.82)	103.40*** (13.25)

Note: \*p<0.1;\*\*p<0.05;\*\*\*p<0.01

## Appendix

Table A1: The Definitions and Coding of the Individual Characteristics in CMDS 2014

Variables	Questions asked in the survey and coding
Origin	<p>“Where is your hukou registered?”</p> <p>1-Beijing; 2-Tianjin; 3-Hebei; 4-Shanxi; 5-Inner Mongolia; 6-Liaoning; 7-Jilin; 8-Heilongjiang; 9-Shanghai; 10-Jiangsu; 11-Zhejiang; 12-Anhui; 13-Fujian; 14-Jiangxi; 15-Shandong; 16-Henan; 17-Hubei; 18-Hunan; 19-Guangdong; 20-Guangxi; 21-Hainan; 22-Chongqing; 23-Sichuan; 24-Guizhou; 25-Yunnan; 26-Tibet; 27-Shaanxi; 28-Gansu; 29-Qinghai; 30-Ningxia; 31-Xinjiang; 32-Xinjiang Bingtuan; 33-Taiwan; 34-Hong Kong; 35-Macao.</p>
Current age	“What is your birth year and birth month?”
Gender	<p>“What is your gender?”</p> <p>1-Male; 2-Female.</p>
Married	<p>“What is your marital status?”</p> <p>1-Single; 2-First marriage; 3-Second marriage; 4-Divorced; 5-Widow or widower.</p>
Current Education	<p>“What is your educational level?”</p> <p>1-No schooling; 2-Primary school; 3-Middle school; 4-High school; 5-College; 6-University; 7-Postgraduate.</p> <p>The educated years is converted from the options: no schooling equals 0; primary school equals 6 years; middle school equals 9 years; high school equals 12 years; college equals 15 years; university equals 16; years; postgraduate equals 19 years.</p>
Number of children	“How many children do you have in total?”
Number of schooling children at migration	The number is calculated from all children’s birth year. 6 to 16 years old are seen as schooling years.
Industry	<p>“Which industry are you working in?”</p> <p>1-Agriculture; 2-Mining; 3-Manufacture; 4-Energy;</p>
Continued on next page	

**Table A1 – continued from previous page**

<b>Variables</b>	<b>Questions asked in the survey and coding</b>
	5-Construction; 6-Wholesale and retail; 7-Transport; 8-Hotels and restaurants; 9-Information technology; 10-Finance; 11-Real estate; 12-Rental and commercial services; 13-Research and technical services; 14-Public facility management; 15-Residential services; 16-Education; 17-Health and social work; 18-Recreational activities; 19-Public administration; 20-International organization. 1 is coded as the primary industry. 2 to 5 are coded as secondary industries. 6 to 20 are coded as tertiary industries.
Income	“What is your income last month from current job or last job (if unemployed)? (in RMB)”
Migration motives	“What is your reason of migration this time?” 1-Work and business; 2-Accompanying migration; 3-Marriage; 4-Forced relocation; 5-Go and live with relatives; 6-Study; 7-Military; 8-Birth; 9-Others.
Years since migration	“How long have lived here since migration?”