北京大学中国经济研究中心
China Center for Economic Research
讨论稿系列
Working Paper Series

E2022016 2022-08-23

Self-Employment and Co-Villager Networks of Internal Migrants in China

Ying Deng and Xiangjun Ma

Abstract

Internal migrants plays a pivotal role in shaping the labor geography in China. However, insufficient attention has been paid to the heterogeneity of internal migrants, particularly regarding their employment types at the destination after migration. This paper studies the impact of co-villager networks on the self-employment propensity of internal migrants in China. We employ a unique nationwide survey data of internal migrants and report comprehensive statistics of internal migrants' self-employment patterns in various dimensions. In particular, we observe large variation of self-employment rates of migrants from different home provinces. Applying the spatial autoregressive (SAR) linear probability model, we find that internal migrants are more likely to become own-account self-employed if their co-villagers are also own-account self-employed and that self-employment decision is not only affected by self-employed co-villagers in the same industry, but also by those in other industries. Moreover, the network has robustly positive effects in certain industries such as wholesale and retail and construction industry. Our results indicate that co-villager networks could help migrants build skills required in a sector and share information, such as tactics of running a business, certain channels of product supply, changes of market demand, etc. This paper fills the literature gap by offering empirical evidence on the co-villager network effect on migrants' self-employment in developing countries.
.

Keywords: Own-Account Self-Employed, Co-Villager Network, SAR, Internal Migrants
JEL Classifications: R23; C21; L26

Self-Employment and Co-Villager Networks of Internal Migrants in China

Ying Deng and Xiangjun Ma*

July, 2022

Abstract

Internal migrants plays a pivotal role in shaping the labor geography in China. However, insufficient attention has been paid to the heterogeneity of internal migrants, particularly regarding their employment types at the destination after migration. This paper studies the impact of co-villager networks on the self-employment propensity of internal migrants in China. We employ a unique nationwide survey data of internal migrants and report comprehensive statistics of internal migrants' self-employment patterns in various dimensions. In particular, we observe large variation of self-employment rates of migrants from different home provinces. Applying the spatial autoregressive (SAR) linear probability model, we find that internal migrants are more likely to become own-account self-employed if their co-villagers are also own-account self-employed and that self-employment decision is not only affected by self-employed co-villagers in the same industry, but also by those in other industries. Moreover, the network has robustly positive effects in certain industries such as wholesale and retail and construction industry. Our results indicate that co-villager networks could help migrants build skills required in a sector and share information, such as tactics of running a business, certain channels of product supply, changes of market demand, etc. This paper fills the literature gap by offering empirical evidence on the co-villager network effect on migrants' self-employment in developing countries.

.

Keywords: *Own-Account Self-Employed, Co-Villager Network, SAR, Internal Migrants*
JEL Classifications: *R23; C21; L26*

*Ying Deng: University of International Business and Economics, 10 Huixin East Street, Chaoyang District, Beijing, China, 100029, ydeng.econ@gmail.com; Xiangjun Ma: University of International Business and Economics, xm2e@virginia.edu. All errors are the responsibility of the authors.

1 Introduction

This paper studies the impact of co-villager networks on the self-employment propensity of internal migrants in China. Prior research has provided some empirical evidence that ethnic networks play a significant role in immigrants' probability of becoming self-employed in developed host countries such as the U.S., Spain, Sweden, and so forth. However, how the network of home origin affects a migrant's self-employment activities in developing countries has rarely been explored.

Existing immigration literature has demonstrated that ethnic networks provide self-employed immigrants with information about markets and institutional conditions; relationships with suppliers, customers, and potential employees; industry specific skills and know-how; knowledge on how to start or take over a business and on related legal and tax-related issues; experience in business methods; activities of interest; acquisition of human and financial capital — all of which are necessary for running a business (Martin-Montaner et al., 2018; Andersson & Hammarstedt, 2012). Borjas (1986) demonstrates a strong, positive impact of assimilation on self-employment rate among immigrants in the U.S. Martin-Montaner et al. (2018) also show that ethnic networks in Spain enhance immigrants' self-employment propensities. However, Andersson & Hammarstedt (2012)'s findings of immigrants from Middle East countries living in Sweden indicate that ethnic networks and enclaves may hinder immigrants' self-employment, because of the increased competition for customers among self-employed immigrants as the network size increases. Another type of intra-group competition which may discourage self-employment is that networks can foster the recruitment of co-national individuals as employees (Martin-Montaner et al., 2018). With the diverse findings in developed countries, it is worthwhile and informative to study the network effect on migrants' self-employment in developing countries.

As the largest developing country, China has the largest scale of internal migrants in the world. Based on the statistics of the National Bureau of Statistics (NBS) of China, there were approximately 385 million internal migrants in China as of 2021, increased by almost 70 percent from a decade ago and accounting for around 27 percent of total population.[1] Because of the Hukou system of China, labor mobility within the nation is not as free as that in other countries.[2] The geographic segregation of citizen identities, as well as the rural and urban Hukou difference, lead to the labor-market segregation, which makes the internal migration in China comparable to international immigration in other nations. Internal migration plays a pivotal role in shaping the labor geography in China.

Although the questions of why people migrate (i.e. migration incentive) and where they migrate to (i.e. migration destination) have been extensively studied (Zhu 2002; Lin et al., 2004; Park 2008), the issue of migrants' employment types at the destination after migration has received much less attention. Self-employment plays a vital part of an economy. First, it contributes

[1] Source: National Bureau of Statistics of China (http://www.stats.gov.cn/tjsj/sjjd/202201/t20220118_1826538.html; http://www.stats.gov.cn/tjsj/zxfb/202105/t20210510_1817176.html)

[2] The Hukou system is a household registration system in China originated in 1951 (Zhao, 2004). Hukou is an official document issued by the Chinese government, certifying that the holder is a legal resident of a particular area. Before the large-scale reform initiated in 1979, labor mobility in China was completely restricted due to the Hukou system, which tied people's residence and work to the place of their Hukou (i.e., home origin). While different provinces have adopted varying policies to gradually allow some degree of mobility after 1979 (for instance, people can now work outside their home province of Hukou), the fundamental Hukou system of residence continues to be in place, and labor mobility remains highly restricted (Cai et al., 2008). In particular, people from rural areas hold rural Hukou, while people from urban areas hold urban Hukou.

to job creation and economic growth, and is associated with upward economic mobility for low-skill workers (Fairlie & Lofstrom, 2015). Employer discrimination against immigrants in the labor market leads to limited upward mobility and pushes them to self-employment (Mata and Pendakur, 1999; Raijman & Tienda, 2000). Second, self-employment facilitates job flexibility and can especially offer more flexible working hours and effort for women (Carr, 1996; Wellington, 2006; Gurley-Calvez et al., 2009). Third, entreprenership among self-employed skilled labor has strong spillover effects in inspiring technology and innovation, fostering employment for others, and enhancing product/service diversities (Fairlie & Lofstrom, 2015). To study the self-employment issue for Chinese internal migrants is particularly important because self-employed migrants make tremendous contribution to the urban economy especially in service industries such as wholesale and retail, accommodation and catering, transportation, construction, etc.[3]

Our study on the co-villager network effect is motivated by the large variation of self-employment rates of migrants from different origin or home provinces. According to our data in 2014 (i.e., the most recent year of the sample), there are 38.4 percent self-employed (own-account plus employer) migrants and 60.0 percent wage workers in the nation, while for migrants from Zhejiang province, the self-employment ratio is 71.1 percent, almost twice as the national level (see Figure 1A & 3A). Fujian is another province from which migrants are popularized by the self-employed type, with a percentage of 67.3. In comparison, there are only 10.6 percent self-employed migrants for Yunnan province. Further, the home cluster of self-employment is even more notable in the large self-employed industries. As we could also see from Columns (2), (8), (11), and (15) of Table 1A, among all own-account self-employed migrants, 5.1 percent are from Zhejiang province, while there are 8.4 percent Zhejiang migrants among all migrants of this type in the Wholesale and Retail industry. Similarly, the proportion of Anhui own-account self-employed migrants in the full sample is 12.2 percent, ranking highest among all 31 provinces, while the percentage is 16.6 percent, even higher, in the Lodging and Catering industry. Sichuan migrants make up 22.4 percent in the Construction industry, but only 10.4 percent in the full sample. (delete the industry statistics?)[4]

What drives the self-employment difference across home provinces? The high self-employment ratios of Zhejiang and Fujian migrants seem to be consistent with the common belief that these provinces are featured in co-villagers' strong mutual-help networks in doing business as well as people's entrepreneurship virtue and ambition. Indeed, two possible explanations emerge in theory. The first one is the co-villager network, the reasoning of which is similar to the ethnic network of immigrants as aforementioned. For U.S. immigrants, Kerr and Mandorff (2015) demonstrate the network effect on occupational concentration of self-employed immigrants across ethnic groups, such as Korean dry cleaners and Indian motel owners, in a theoretical model, which emphasizes social networks' cost reduction mechanism in acquiring sector-specific skills for entrepreneurship. In China, co-villager network is one of the most important social capital which deeply influence internal migrants' labor market activities in many aspects. Previous work on co-villager networks finds

[3]Most existing studies on internal migrants' self-employment in China focus on how the labor market discrimination against the rural-Hukou identity limits migrants' job opportunities so that they are more likely to enter into self-employment (Meng and Zhang, 2001; Gagnon et al., 2012).)

[4]Similar employer pattern of industry-level home clusters can be found in Columns (2), (8), (11), and (15) of Table 1B, which present the "absolute proportion" of self-employed migrants from each home province. Note that Table 1A & 1B only list the top 10 home provinces ranked by the "relative proportion," which substracts the share of migrants from a home province among all migrants in the nation from the absolute proportion. More explanation will be provided in Section 2.2.

that experienced migrants have a positive and significant effect on subsequent migration (Zhao, 2003; Chen et al., 2010). For self-employment, a close study is Zhang and Zhao (2011)'s two-stage least squares (2SLS) estimation on social-family network effects. The second explanation shall be migrants' common characteristics and culture shaped over the histories of their home origins, such as the entrepreneurship tradition. Hammarstedt and Shukur (2009) and Yuengert (1995) verify the home country self-employment hypothesis that explains the high self-employment rates among immigrants in a country and interpret experience in the self-employment sector as a form of sector-specific human capital. Certain comparative advantage in the production of goods or services (e.g., food or restaurant service) that is specific to people from a given ethnic enclave also boosts migrants' self-employment propensity (Martin-Montaner et al., 2018). As the second explanation has been massively studied in the literature, this paper emphasizes the network channel and use a spatial autoregressive (SAR) model to identify the co-villager network effect on migrants' self-employment decisions.

We employ a large nationwide survey data of internal migrants in China from 2011 to 2014, which has comprehensive demographic and work variables and covers around 330 host cities all over the 31 provinces of China. We first present extensive statistics of internal migrants' self-employment patterns and then apply the SAR model, which has been widely used in empirical analyses of social networks, [5] to test whether a migrant's propensity to becoming self-employed is higher if most of her/his co-villagers are self-employed. Existing empirical work usually measure networks by multiplying the size of the ethnic group and the average self-employment rate of the group (Bertrand et al., 2000; Martin-Montaner et al., 2018; Andersson & Hammarstedt, 2012). Our SAR methodology offers a more direct examination on the network effect. To deal with the endogeneity concern that the self-employment decisions of the co-villagers in the network can in turn be influenced by the objective migrant's self-employment decision in a symmetric way, we follow the spatial two-stage least squares (S2SLS) estimates proposed by Kelejian and Prucha (1998).

Our benchmark results confirm the positive effect of co-villager networks on the own-account type of self-employment. Similar to Martin-Montaner et al. (2018)'s finding in Spain, the network effect on employers is insignificant. For own-account migrants, their self-employment decision is not only affected by co-villagers in the same industry, but also by those in other industries, indicating that co-villager network provides general support on starting a business, while the industry specific assistance is just part of it. In addition, the network has the robustly positive effects in certain industries such as wholesale and retail and construction industry. Furthermore, we use a difference-in-difference (DID) specification and show that the network effect is stronger if the individual hangs out with co-villagers most in her/his life. Similarly, the other set of DID results also show that the network effect is stronger if the individual has stayed longer in the host city. These findings can be viewed as the mechanism proof of the co-villager network effect, because (1) only when a migrant is truly utilizing her/his co-villager network, and (2) only when a migrant has stayed a sufficient amount of years in the host city, she/he can have a more effective and influential network. Lastly, our subsample analyses demonstrate that the network effects are larger

[5] LeSage and Pace (2009) have a detailed introduction of the SAR model. Its applications are widespread from studying spatial interactions at the macro level (countries, cities, etc.) to investigating social interactions at the micro level (households, individuals, etc.). Lin (2010) applies the SAR model to identify the peer effects in students' academic achievements. Baltagi and Yen (2014) allow spatial correlation among neighboring hospitals and estimate the effects of externalities generated by competition and knowledge spillovers on hospital treatment rates.

among males, rural migrants, migrants with very low or very high education, and migrants residing in larger host cities.

In sum, to recognize the home clusters of self-employment is important for researchers to understand the self-employment distribution of internal migrants in China. This paper fills the literature gap by offering empirical evidence on the co-villager network effect on migrants' self-employment in developing countries. The remainder of the paper is organized as follows. Section 2 reports the statistics of self-employment patterns of internal migrants. Section 3 and 4 present the SAR model and data source. Section 5 discusses the basic empirical results and performs the heterogeneity analysis. Section 6 concludes.

2 Self-Employment Patterns and Home Clusters of Internal Migrants in China

2.1 General Self-Employed Patterns

Based on our data in 2014 (i.e., the most recent year of the sample), we report and analyze the statistics of internal migrants' self-employment patterns in various dimensions in Figure 1-3.

Figure 1A shows that the majority of migrants are wage workers or employees (around 60 percent), followed by own-account self-employed migrants (30 percent) and employers (nine percent). The rest are other employment types who are mainly house helpers working in their families. In comparison, the overall self-employment rate in China, regardless local residents or internal migrants, is 47.8 percent in 2014, based on the World Bank statistics,[6] which is higher than the rate of migrants revealed in our data. Though migrants are likely to be pushed to self-employment due to their disadvantaged position in education and Hukou status when competing with natives for wage jobs, the result here is not surprising because a substantial amount of internal migrants work as employees in manufacturing when China was playing the "world factory" role during that time span.

A migrant's entry into an employment type and an industry is often a joint decision because some industries are by nature more self-employed based. Therefore, we then show the self-employment composition for major industries. We first display the industry distribution of internal migrants in Figure 2A and identify the top five large industries as Manufacturing, Wholesale and Retail, Lodging and Catering, Other Industries, and Construction, which have significant larger shares than the rest of the industries and constitute over 80 percent of total employment in the sample. As we know little about the exact content of "Other Industries," we only report the employment composition in the same pie graph, Figure 1B, for each of the other four large industries. We observe large variation in employment types across industries: while self-employment is very rare in Manufacturing, own-account self-employed (61.9 percent) and employer (18.9 percent) are the dominant types in Wholesale and Retail; the percentage of self-employment is also high (48.9 percent as a whole) in Lodging and Catering, but the wage worker type is the majority in Construction (68.7 percent). In fact, if we rank all industries by the self-employment ratio, "Wholesale and Retail", "Lodging and Catering", and "Construction" are also among the top half industries (see Figure 2B). Besides, "Agriculture, Forestry, Husbandry, and Fishing" and "Transportation,

[6] Data source: International Labour Organization, ILOSTAT database (https://data.worldbank.org/indicator/SL.EMP.SELF.ZS).

Storage, and Communication" also have high ratios of own-account self-employed migrants due to the nature of these types of work. In our empirical analysis, we will specifically test the network effects on self-employment in these four largest industries, separately.

Next, we display the self-employment rates in more dimensions in Figure 3. Figure 3A shows the percentages of own-account self-employed and employer types for each home province. Among the top 10 provinces with highest own-account percentages, there are seven southern provinces, two western provinces, and only one northern province, which exhibits a clear south-bias style. On the contrary, the bottom three provinces (excluding Beijing and Shanghai) are all southwestern ones.[7] Similarly, among the top five ranked provinces in term of employer rate (excluding Beijing and Shanghai), most of them are southern provinces.

Figure 3B shows the self-employment rates in each education group. For own-account migrants, the rate decreases as the education level increases. Gindling & Newhouse (2014)'s findings in developing countries also show that years of schooling is highest for employers, followed by wage workers, and lowest for own-account workers. Van der Sluis et al. (2008) states that it is easier for highly educated people to find high-wage jobs, which increases the opportunity costs of self-employment. However, the literature review of Simoes et al. (2016) and Van der Sluis et al. (2008) also summarize that theoretical prediction on the overall relationship between schooling and entrepreneurship selection is unclear because educated people can also better identify self-employment opportunities and have greater managerial ability, analytical and communication skills that are needed to run a business; therefore, an inverted U-shape relationship is often found in the literature. In fact, for our data of employer rates, an inverted U-shape relationship with education does exhibit.

Figure 3C tells that the self-employment rates are higher when the migrant is a male, has married, has at least one child living together with her/him in the host city. These results are highly consistent with the literature. Women have a lower propensity to choose self-employment than men as a result of their higher level of risk aversion and more family responsibilities that prevent them from investing time in networking (see a review from Simoes et al, 2016). Married migrants usually have more resources and wealth than single individuals, gain physical assistance as well as emotional support from the spouse for their self-employment activities (Borjas 1986; Bosma et al., 2004), and benefit from sharing the spouse's skills and network of contacts. Families with the presence of children turn to self-employment because they appreciate the flexibility in managing working time and are more motivated to seek for higher expected returns (Dawson et al., 2013). We also observe that migrants with rural Hukou are more likely to become own-account self-employed and less likely to become employers than urban Hukou holders, which may be explained by the labor market discrimination of wage jobs, as well as the low education level, on average, of rural migrants.

Lastly, Figure 3D shows how the self-employment rates change as migrants' age and years of residence in the host city increase. Both own-account and employer rates display an inverse U-shape relationship with both age and years of residence. Simoes et al. (2016) summarize that on the one hand, older people have accumulated more human capital, financial capital, and social

---

7

Beijing and Shanghai are the two largest migrant-receiving provinces/municipalities and the most developed areas in China. The migration incentives of people from Beijing and Shanghai are very different from the majority of Chinese internal migrants so that their self-employment pattern is not comparable with others.

capital (Calvo & Wellisz, 1980; Van Praag & Van Ophem, 1995); on the other hand, older people are more risk averse and are less physically and mentally available to work for long hours and under big stress (Hintermaier and Steinberger, 2005). Clark et al. (2017) summarize that time since arrival is associated with the accumulation of human capital, physical capital, and financial capital, which are all important factors of business establishment; as people stay in the host cities even longer, they also grow older, and the diminishing part of the inverse U-shape curve may be explained by the same reason for that of age.

2.2 Home Province Clusters of Self-Employment across Industries

In addition to the variations in self-employment rates across industries or home provinces shown in Figures 2B & 3A, we also observe home province clusters of self-employment given an industry. For the four largest industries, Manufacturing, Wholesale and Retail, Lodging and Catering, and Construction, as well as all industries as a whole, the "Absolute Proportion" columns of Table 1A present the proportions of own-account migrants of each home province (among all own-account migrants of the nation) in each industry. We can see that migrants from Anhui and Henan make up the highest self-employment proportions for all industries as a whole. However, this may be driven by the high shares of migrants from these home provinces among all migrants in the nation, regardless of the employment type. Therefore, we rely on the *relative proportion* that is obtained by subtracting "the share of migrants from a home province among all migrants in the nation" from the *absolute proportion*. Table 1A reports the top 10 home provinces ranked by the *relative proportion*. As we see now, in general, Zhejiang and Fujian are the two largest provinces with the highest relative own-account proportions, followed by Henan and Hunan. Zhejiang, Fujian, and Hunan also rank the top three in the Wholesale and Retail industry. Anhui, Fujian, Gansu, Chongqing, Shanxi, and Qinghai have the highest relative own-account proportions in Lodging and Catering, while Sichuan and Chongqing have significantly higher relative proportions than others in Construction.[8] Jiangxi, Hubei, and Anhui are the top three provinces in Manufacturing, though we have known from Figure 1B that there are not many self-employed migrants in this industry.

Table 1B report the same statistics for employers which display slightly different home province clusters. Besides Zhejiang, Fujian, and Hunan, Jiangsu and Guangdong stand out as employer hometowns for all industries as a whole. They are also the top five home provinces of employers in Wholesale and Retail. The top six provinces of own-account self-employed in Lodging and Catering are also the top ones for employers. However, the top three Construction employer provinces become Jiangsu, Sichuan, and Fujian. Jiangxi, Hunan, Anhui, and Zhejiang are the top home provinces in Manufacturing for employer migrants.

The facts revealed in Table 1A & 1B are all consistent with people's traditional impressions on the geographic distribution of self-employed migrants in China. For example, most of the top provinces in Lodging and Catering are well-known for their locally specialized gourmet, meaning that migrants from these provinces have comparative advantage at running restaurants. Even

[8] Note that before Chongqing was designated as one of the four municipalities in China in 1997, it was a city of Sichuan province, which further reinforces the geographic home cluster feature of own-account self-employed migrants in Construction.

though certain home-based comparative advantage might be the key determinant pushing pioneer migrants to self-employment, the co-villager network can well possibly play a crucial role that helps form dynamically the home clusters of self-employment afterwards, which we will test empirically in Section 4.

3 Data and Summary Statistics

3.1 Data Source

Our migrant data comes from the China Migrant Dynamic Survey (CMDS), conducted by the National Population and Family Planning Commission of China. We employ the 2011-2014 data of this annual survey when the most comprehensive individual variables are available.[9] In the four years of our sample, the survey comprises of 128,000 - 200,937 individual migrants who are between ages 16 and 59 and residing in 326 - 337 prefecture-level or county-level host cities in all 31 provinces as well as Xinjiang Production and Construction Corps. In each province, 2,000 - 15,000 individuals are surveyed depending on the population size of the province.[10] The survey includes migrants' demographic information, such as birth year, gender, education level, type of Hukou (rural or urban Hukou), home province (registration province of Hukou), ethnicity, employment status, industry, marriage status, number of children living in the host city and home province, number of years of residence in the host city, whom they hang out most in current life, and so forth. Internal migrants in China could largely be classified into within-province migrants and cross-province migrants. In this paper, we focus on the latter because the CMDS does not ask migrants' home cities and thus only home province information is available, which means the home province will be the same as the host province for within-province migrants and we cannot distinguish a migrant's origin and destination. Based on the 2014 data, 51 percent of total migrants in our sample are engaged in cross-province migration.

As a nationwide data officially collected by the government agency, it is superior to other migrant survey data collected by individual research groups in terms of population scope. For example, another widely used migrant survey data is the Migrant Household Survey (MHS) as part of the Longitudinal Survey on Rural Urban Migration in China (RUMiC), which includes 5,000 migrant households floating from 9 western and central provinces and residing in 15 cities concentrating in Southeast China (Akgüç et al, 2014). Although the MHS data has many advantages in multiple control groups to investigate the effects of rural–urban migration (Kong, 2010), the CMDS covering all over the nation allows us to examine the impact of co-villager networks of a full range of locations in China on the migrant employment pattern, Moreover, the CMDS can better serve the purpose of our paper compared with the population census data of China because the former covers more recent consecutive years.

3.2 Summary Statistics of Variables

[9] A long time series of this data up to 2017 is publicly available. However, the most comprehensive individual variables are only available in 2011-2014, which are confidential.

[10] The survey was conducted administratively by government agents at communities. Random migrants were drawn from representative communities each year. Therefore, this is a repeated cross-sectional data.

Our key outcome variables are the two self-employment dummies: (1) $Ownaccount_i$, equal to one if the migrant is an own-account self-employed individual, and zero if she/he is a wage-worker or house helper (i.e. other employment type in Figure 1), and (2) $Employer_i$, equal to one if the migrant is an employer, and zero if she/he is a wage-worker or house helper.[11]

We define a co-villager network as people from the same home province and currently residing in the same host city.[12] In spite of the large population of home provinces in China, the average size of the network restricted in a given host city based on the survey data is not very big. In our 2014 sample with 102,403 individuals, the median size of networks is six co-villagers. The mean and standard deviation of the network size are 26 and 89, respectively.

Other control variables include a migrant's gender (a dummy equal to one if the individual is male and zero if she is female), education level (a dummy equal to one if the individual has high school or above education and zero if otherwise), marriage status (a dummy equal to one if the individual has married and zero if she/he has never married), whether a migrant has any children living together (a dummy equal to one if the individual has at least one child living with her in the host city and zero otherwise), age, square of age, years she/he has stayed in the host city, square of years she/he has stayed in the host city, Hukou (a dummy equal to one if the individual has rural Hukou and zero if she/he has urban Hukou), ethnicity (a dummy equal to one if the individual is Han and zero otherwise). In our channel analysis, another important variable is the $Hangout_i$ dummy, which is equal to one if the migrant hangs out most with a co-villager friend, and zero if she/he hangs out most with local people or does not hang out with people much.

Table 2 presents summary statistics for all variables. In the full sample, the means of $Ownaccount_i$ and $Employer_i$ are 0.323 and 0.127. Note that the $Hangout_i$ dummy is only available in 2011 and 2012. There are also a lot of missing values for years a migrant has stayed in the host city in 2012.

4 Methodology

As aforementioned in the introduction, co-villager networks help migrants achieve cost reduction in acquiring sector-specific skills and can provide migrants with necessary resources in market information, suppliers, customers, human and financial capital, knowledge and experience sharing, etc. so as to facilitate their self-employment activities. Although networks may have negative impact on self-employment due to the increased competition for customers among the self-employed migrant community and the higher probability of recruiting co-villagers as employees, the strong pattern of home province clusters revealed in the statistics of self-employed migrants in China, as shown in Section 2, suggest the positive co-villager network effects on self-employment. Therefore, we empirically examine the following hypothesis:

Hypothesis: More self-employed migrants in an individual migrant's co-villager network increase her/his propensity of being self-employed.

We estimate an SAR linear probability model given by

[11] The unemployed migrants and those in school are dropped out of the sample.

[12] Though we term the network as "co-villager network" in our paper, we do not know a migrant's home city, county, or village. Home province is the only home origin information available in the data.

$$Selfemp_i = \lambda \sum_{j=1}^{n} w_{ij} \cdot Selfemp_j + X_i\beta + Home_h + City_c + Year_t + \varepsilon_i, \quad i = 1, 2, ..., n, \quad (1)$$

where the outcome variable $Selfemp_i$ can be $Ownaccount_i$ or $Employer_i$. The spatial lagged dependent variable $\sum_{j=1}^{n} w_{ij} \cdot Selfemp_j$ is a weighted average of the self-employment dummy of all other migrants in the co-villager network of migrant i. We specify the spatial weight matrix based on migrant i's co-villager network, which is defined as people flowing out of the same home province and residing in the same host city as migrant i. Only people within the network will make an impact on each other and we assume that each migrant is equally affected by all others in the co-villager network. Therefore, each co-villager has an identical influential weight which is the inverse of the number of co-villagers in the network; other migrants who are not co-villagers have zero influential weight. For example, if there are 40 migrants from Sichuan province working in Beijing, then for each migrant, her/his Sichuan-Beijing network consists of 39 people and each person is assigned a weight of 1/39 in order to satisfy the row-normalized condition of the spatial weight matrix. As a result, the spatial lagged dependent variable is the weighted average of the self-employment decision of these 39 migrants in the network with the same weight of 1/39. Other migrants in the sample who do not belong to the Sichuan-Beijing network are assigned zero weights. The diagonal elements are also set to zero. Therefore, the spatial weight matrix W is an $n \times n$ predetermined, row-normalized, symmetric, block-diagonal, sparse matrix, with a typical element w_{ij} representing the weight of migrant j in i's network. Hence, $w_{ii} = 0$ and $\sum_{j=1}^{n} w_{ij} = 1$ for $i = 1, 2, ..., n$. The coefficient λ captures the network effects on the probability of becoming self-employed.

The control variables have been described in Section 3.2, which are denoted as X_i, a $1 \times k$ vector of exogenous characteristics of migrant i, in Equation (1). We also control for the home province ($Home_h$), host city ($City_c$), and year ($Year_t$) fixed effects. ε_i is an independent error with mean zero variance $\tilde{\sigma}_i^2$. β is a $k \times 1$ vector of coefficients of these exogenous regressors.

An endogeneity issue arises as the self-employment decisions of the co-villagers in the network are also influenced by the self-employment decision of the objective migrant in a symmetric way. In other words, $\sum_{j=1}^{n} w_{ij} \cdot Selfemp_j$ is correlated with the error term ε_i. To achieve consistent estimates, we follow the S2SLS estimates proposed by Kelejian and Prucha (1998), who suggest using all the exogenous variables to construct a set of instruments for the endogenous spatial lagged dependent variable. Define $X = (X_1', X_2', ..., X_n')'$ as an $n \times k$ matrix of all exogenous regressors, and then the instrument set is (X, WX).[13]

5 Empirical Results

5.1 Basic Results of co-villager Network Effects on Own-Account Self-Employed Propensity

[13] See Lee (2003) and Kelejian, Prucha and Yuzefovich (2004) for the S2SLS estimation with different sets of instruments.

Table 3A presents the basic estimation results of the SAR linear probability for own-account self-employed migrants. As specified in Section 4, we use $Z = (X, WX)$ as a set of instruments to deal with the endogeneity problem resulting from the spatial lagged dependent variable. X represents a set of exogenous variables as described in Section 3.2. Column (1) and (2) present the basic results while Column (3) and (4) present the results when constructing weights for network co-villagers working in the same industry. The latter captures the network effects arising from the same-industry influence, i.e., whether self-employed co-villagers in the same industry promote an individual's self-employed decision more. In Columns (2) and (4) we add the fixed effects of the home province, current residential city, industry, and year.

Both the basic and the same-industry results show significant positive coefficient estimates of the spatial lagged dependent term. This implies that a migrant's probability of entering own-account self-employment increases with the own-account self-employment decisions of other migrants in her co-villager network. Taking Column (2) as the baseline outputs for the SAR linear probability model with the most strict controls, the result suggests that when the percentage of migrants in the co-villager network accepting jobs without contracts increases by 10%, the probability of the targeting migrant becoming self-employed increases by about 1.2%. The corresponding marginal effect is 0.6% when only considering the influence from co-villagers from the same industry. An alternative explanation for this positive coefficient can be: people from the same origins may share common culture such as self-employment tradition. However, our home province fixed effects control for such factor and serves to isolate the network effect.

Our estimates also imply that male, low-educated, married, older, and rural-Hukou migrants, as well as the migrants who have kids or have stayed longer in the host cities, are more likely to become own-account self-employed.

5.2 Basic Results of co-villager Network Effects on Employer Propensity

As shown in Table 8B, after controlling for fixed effects, the network effects on migrants' decision of becoming employers are not robustly significant, either for the same-industry co-villagers or all co-villagers regardless of industries. This result is consistent with Martin-Montaner et al. (2018)'s finding for migrants in Spain. There might be two possible explanations. First, on the one hand, more employers in the co-villager network may enhance the chance that the targeting individual develops to an employer with the help of her/his co-villagers; on the other hand, more employers in the network may provide more wage worker opportunities for the targeting individual. Thus the negative effect may offset the positive effect. Second, there are relatively fewer employers in the internal migrant data of China and the chance for a targeting individual to know co-villager employers is lower than that in the own-account self-employed case. Therefore, our analysis will focus on the own-account self-employment case in the rest of the paper.

5.3 Co-villager Network Effects on Own-Account Self-Employed Propensity in Major Industries

We also investigate network effects in different industries, because it is easier to become self-employed in certain industries by nature. The top 4 industries in the survey are manufacturing, wholesale and retail, lodging and catering, as well as construction. Table 4 reports the estimates using the same control variables and fixed effects as in the basic model, but excluding industry fixed effects. As we can see, the network has the robustly positive effects in the wholesale and retail and construction industry, which are significant for both general network and same-industry network. This shows that migrants working in these industries are more likely to find jobs through networks or rely on networks to gain benefits. However, manufacturing and lodging and catering only have positive network effect for the same-industry network. For manufacturing, to become self-employed professionals may require specific skills that can only be assisted by people from the same industry. For lodging and catering, it may also require certain upstream suppliers or channels, which may also be likely to be provided by people in the same industry.

5.4 Mechanism Investigation

We caution that people from the same home province in the same host city in our survey data do not necessarily know each other.[14] Nevertheless, since the data is collected in a random way, the co-villager network defined in the data is representative of the true network of the population, which means the size of the networks in the sample is also proportional to the size of the true co-villager networks. In other words, we assume that if there are more people from Sichuan province working in Beijing in the true population, for example, we would see more Sichuan people in Beijing in our data. Further, we also assume that if there are more people in the defined network, there will be more interconnection occurring among them in the network.

We are aware that the second assumption is a little strong as people from different regions may have different co-villager culture and there may not be sufficient job-related interconnection among people in some co-villager networks. As a result, we perform channel regressions using a DID technique by interacting the network variable ($\sum_{j=1}^{n} w_{ij} \cdot Selfemp_j$) with two variables: (1) a dummy equal to 1 if the individual hangs out most with co-villagers ($Hangout_i$) or (2) number of years she/he has stayed in the host city. We take $Hangout_i$ as a channel variable because only when the migrant hangs out with co-villagers, the network influence can take effect. Similarly, when the migrant has stayed in the host city for longer time, she/he can have a more effective and influential network and there can be more chances for the migrant to be affected by the network. Positive coefficients of the interaction terms would confirm the channels through which the network of self-employed co-villagers could promote more self-employment. Since the "Years of Stay" variable may have non-linear effects, we also add the interaction term of $\sum_{j=1}^{n} w_{ij} \cdot Selfemp_j$ and the square of "Years of Stay".

The results in Table 5 show that hanging out more with co-villagers and staying longer in the host city do increase the network effect on migrants' self-employment propensity. Such channel analysis verifies the network effects indirectly.

5.5 Heterogeneity Analysis and Robustness Check

We then examine the co-villager network effects in the following subsamples by reproducing the benchmark estimation with all controls and fixed effects as in the basic regressions: (1) males

[14] There are other applications of the spatial methodology in the literature, where people in the same group do not necessarily know each other. For instance, Lin et al. (2006) use occupation and township as a way to identify connectivity between individuals in a study of national identity formation in Taiwan using spatial regressions. In this paper, people may not know each other but they belong to the same occupational group or township group.

versus females, (2) low education versus high education, (3) rural Hukou versus urban Hukou, and (4) host cities of three levels of sizes. The estimates are all significant in all the subsamples but with different magnitude. It is shown in Table 6 that the self-employment decisions of male migrants and rural migrants are more influenced by co-villager networks than females and urban migrants. Migrants with very low education or very high education are more likely to be influenced by networks, which may be due to the fact that middle-educated migrants are more likely to become wage workers. The network effect is stronger among internal migrants in the larger host cities, and strongest in the four largest host cities - Beijing, Shanghai, Guangzhou, and Shenzhen.

6 Conclusion

This paper employs a comprehensive data set from a nationwide survey to study how co-villager networks affect internal migrants' self-employment propensity in China. Surrounded by a network with a large amount of self-employed co-villagers, the migrant may be motivated to start her/his own business and receive assistance in information about markets and institutional conditions; relationships with suppliers, customers, and potential employees; industry specific skills and know-how; knowledge on how to start or take over a business and on related legal and tax-related issues; experience in business methods; activities of interest; acquisition of human and financial capital. We first present comprehensive statistics of internal migrants' self-employment patterns in various dimensions. We then apply the SAR linear probability model and verify that internal migrants are more likely to become own-account self-employed if their co-villagers are also own-account self-employed and that self-employment decision is not only affected by self-employed co-villagers in the same industry, but also by those in other industries. Moreover, the network has robustly positive effects in certain industries such as wholesale and retail and construction industry. Furthermore, we use a DID specification and show that the network effect is stronger if the individual hangs out with co-villagers most in her/his life or has stayed longer in the host city. Lastly, our subsample analyses demonstrate that the network effects are larger among males, rural migrants, migrants with very low or very high education, and migrants residing in larger host cities. This paper fills the literature gap by offering empirical evidence on the co-villager network effect on migrants' self-employment in developing countries.

Appendix

Figure 1: Composition of Employment Types of Internal Migrants in China (2014)

Figure 1A: Composition of Employment Types of All Migrants

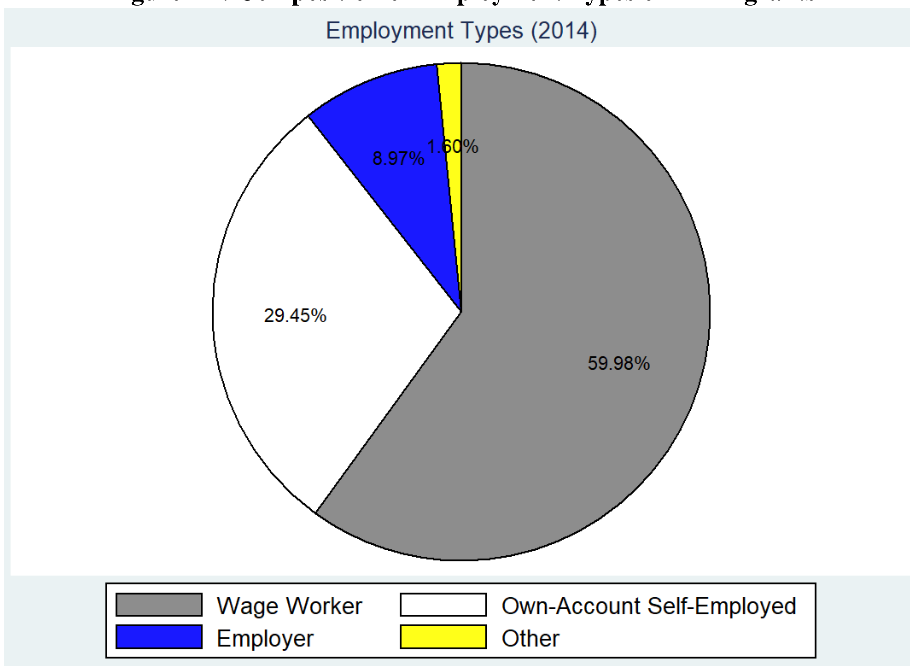

Figure 1B: Proportion of Self-Employed Migrants for Each Industry

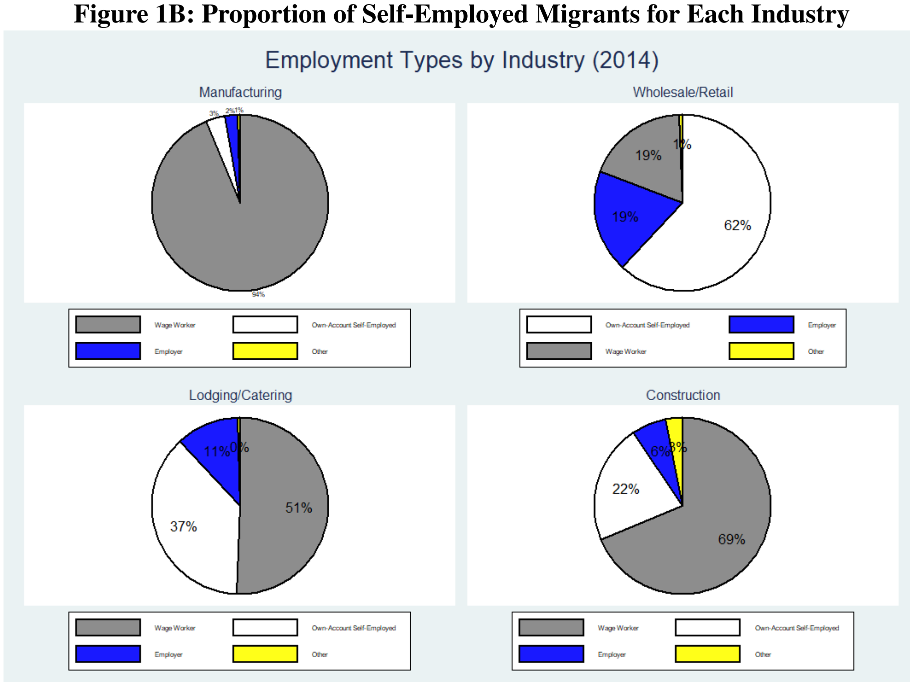

Figure 2: Industry Distribution and Self-Employment Rates (2014)

Figure 2A: Industry Distribution of All Migrants

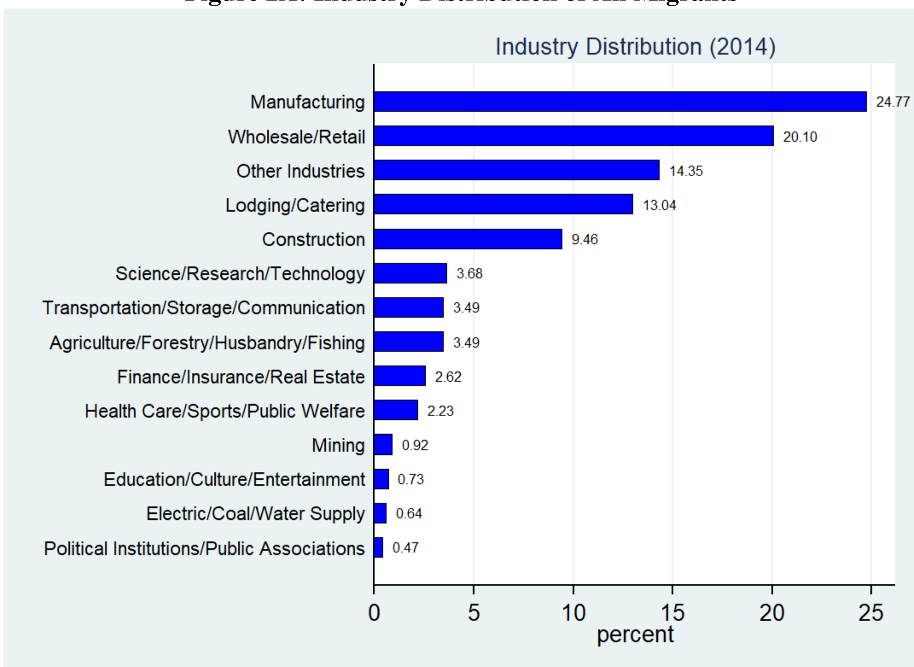

Figure 2B: Self-Employed Rates of Migrants for Each Industry

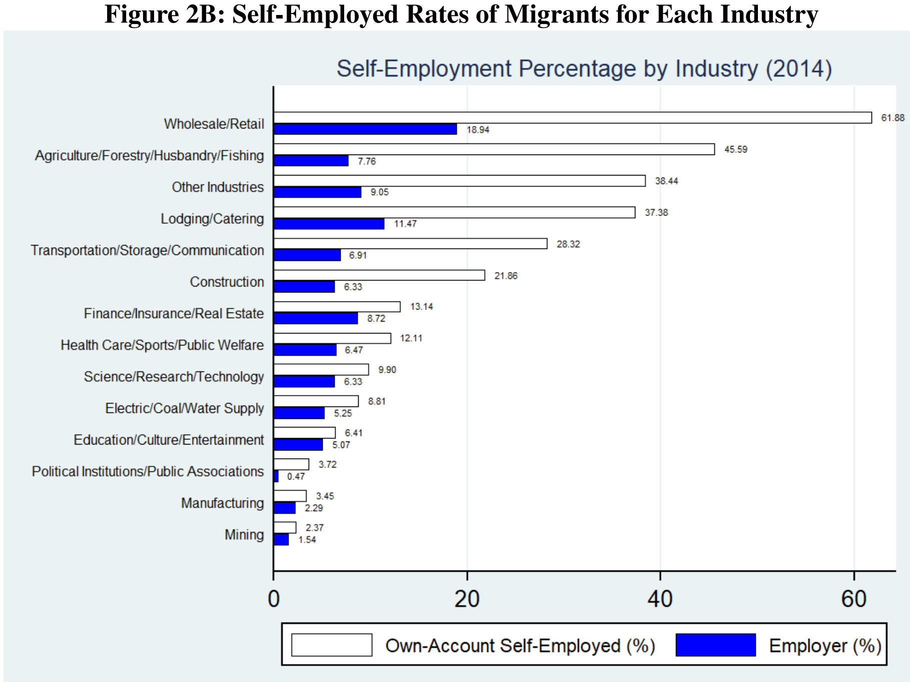

Figure 3: Self-Employment Pattern of Various Dimensions (2014)

Figure 3A: Self-Employment Rates by Home Province

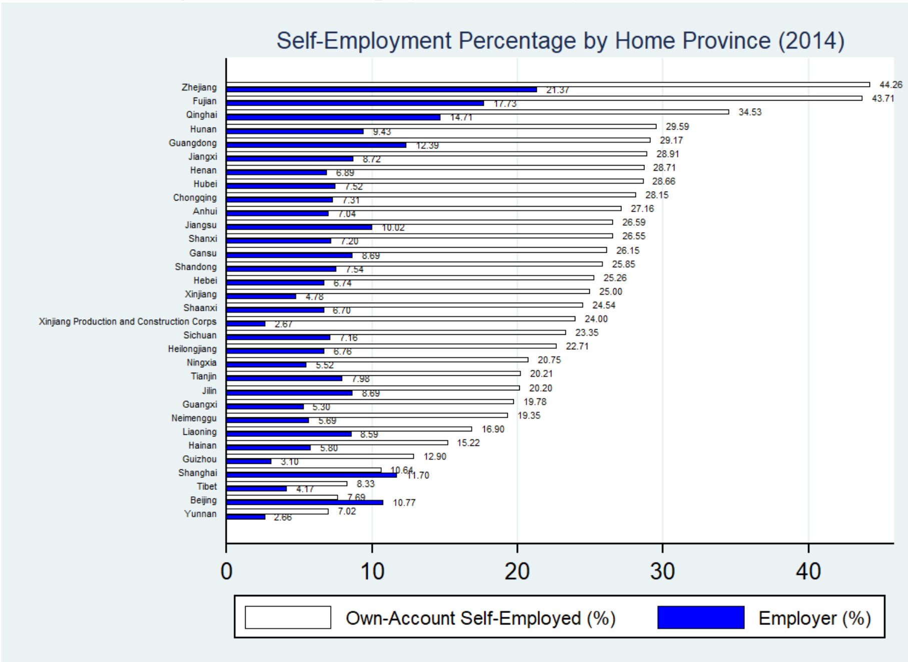

Figure 3B: Self-Employment Rates by Education Level

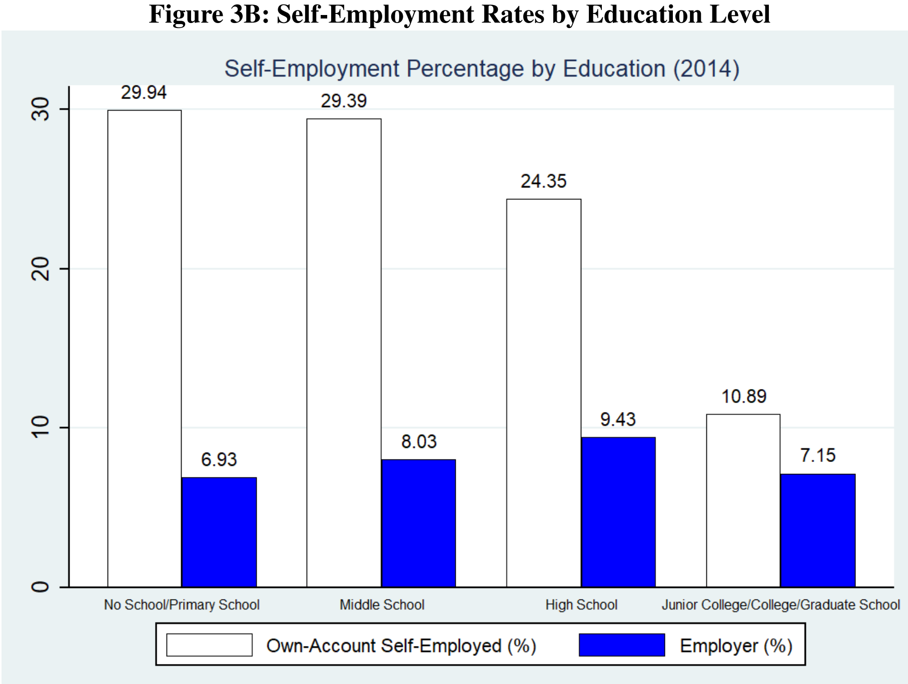

Figure 3C: Self-Employment Rates by Gender, Marriage Status, Children Status, and Hukou Status

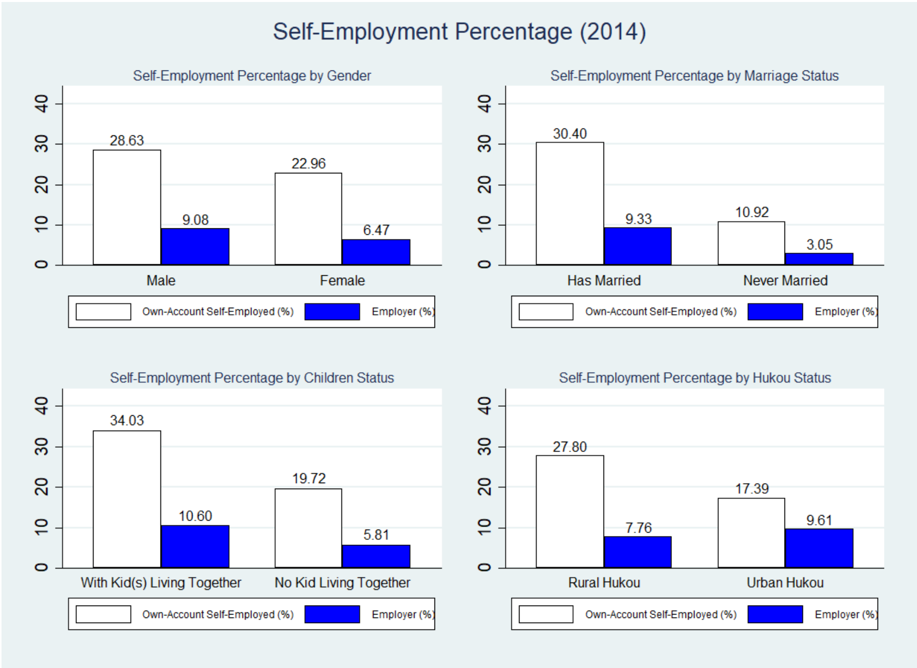

Figure 3D: Self-Employment Rates by Age and Years of Residence in Host Cities

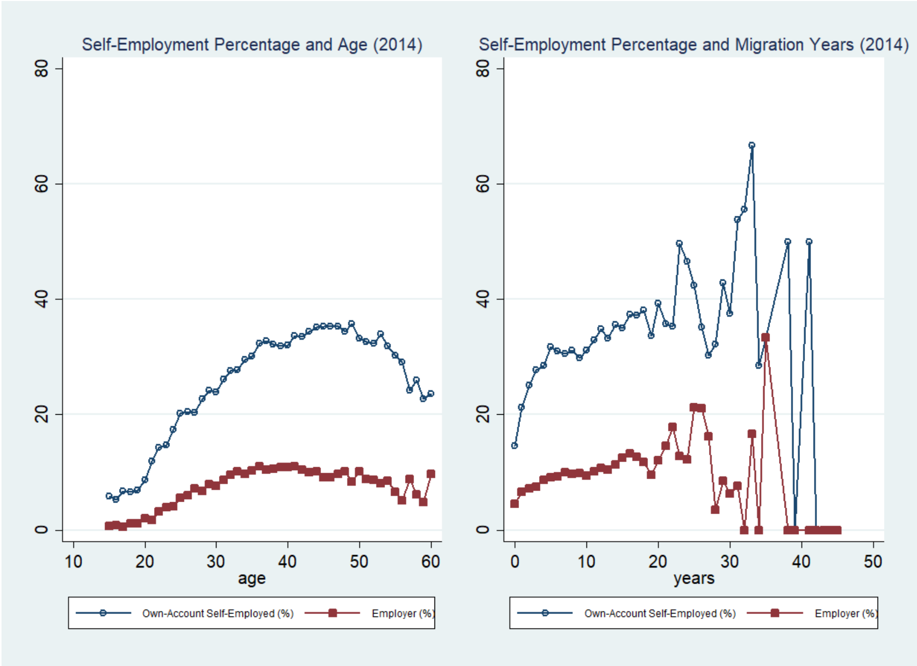

Table 1: Relative Home Province Proportions of Self-Employed Migrants

Table 1A: Relative Home Province Proportions of Own-Account Self-Employed Migrants

Rank	**All Industries**			**Manufacturing**			**Wholesale and Retail**			**Lodging and Catering**			**Construction**		
	Home Province	Absolute Proportion	Relative Proportion	Home Province	Absolute Proportion	Relative Proportion	Home Province	Absolute Proportion	Relative Proportion	Home Province	Absolute Proportion	Relative Proportion	Home Province	Absolute Proportion	Relative Proportion
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
1	Zhejiang	5.08%	2.06%	Jiangxi	11.40%	6.06%	Zhejiang	8.36%	5.34%	Anhui	16.57%	4.79%	Sichuan	22.35%	11.07%
2	Fujian	4.31%	1.71%	Hubei	10.24%	4.70%	Fujian	6.04%	3.45%	Fujian	5.27%	2.68%	Chongqing	5.92%	2.40%
3	Henan	11.86%	0.99%	Anhui	16.39%	4.61%	Hunan	8.77%	2.60%	Gansu	5.65%	1.80%	Jiangxi	5.92%	0.58%
4	Hunan	6.94%	0.77%	Hunan	8.45%	2.28%	Hubei	6.72%	1.18%	Chongqing	5.27%	1.75%	Jiangsu	3.65%	0.56%
5	Jiangxi	5.86%	0.53%	Jiangsu	3.97%	0.88%	Hebei	5.52%	0.92%	Shanxi	2.80%	1.41%	Hubei	5.97%	0.43%
6	Hubei	6.04%	0.49%	Fujian	3.33%	0.74%	Jiangxi	6.05%	0.72%	Qinghai	1.39%	1.06%	Fujian	2.80%	0.21%
7	Anhui	12.15%	0.38%	Guangxi	3.07%	0.66%	Shandong	6.40%	0.71%	Jiangxi	6.12%	0.79%	Hunan	6.23%	0.06%
8	Chongqing	3.76%	0.24%	Guizhou	3.97%	0.41%	Jiangsu	3.65%	0.56%	Shaanxi	3.05%	0.56%	Tianjin	0.21%	0.03%
9	Guangdong	1.23%	0.12%	Hainan	0.13%	-0.01%	Guangdong	1.59%	0.48%	Ningxia	0.49%	0.05%	Guangdong	1.11%	0.00%
10	Qinghai	0.43%	0.10%	Yunnan	1.41%	-0.14%	Henan	11.10%	0.23%	Xinjiang	0.65%	0.02%	Hainan	0.11%	-0.03%

Table 1B: Relative Home Province Proportions of Employer Migrants

Rank	**All Industries**			**Manufacturing**			**Wholesale and Retail**			**Lodging and Catering**			**Construction**		
	Home Province	Absolute Proportion	Relative Proportion	Home Province	Absolute Proportion	Relative Proportion	Home Province	Absolute Proportion	Relative Proportion	Home Province	Absolute Proportion	Relative Proportion	Home Province	Absolute Proportion	Relative Proportion
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
1	Zhejiang	8.05%	5.03%	Jiangxi	11.15%	5.82%	Zhejiang	12.69%	9.67%	Fujian	5.56%	2.96%	Jiangsu	8.03%	4.94%
2	Fujian	5.74%	3.14%	Hunan	11.15%	4.98%	Fujian	7.61%	5.01%	Gansu	6.51%	2.65%	Sichuan	15.88%	4.60%
3	Hunan	7.26%	1.09%	Anhui	16.54%	4.76%	Hunan	7.87%	1.70%	Anhui	13.96%	2.19%	Fujian	5.47%	2.88%
4	Jiangsu	3.86%	0.77%	Zhejiang	6.15%	3.13%	Jiangsu	4.48%	1.39%	Shanxi	3.14%	1.75%	Jiangxi	6.57%	1.24%
5	Guangdong	1.72%	0.61%	Guangxi	4.04%	1.62%	Guangdong	2.04%	0.93%	Chongqing	5.04%	1.53%	Henan	12.04%	1.17%
6	Jiangxi	5.80%	0.46%	Hubei	7.12%	1.57%	Hubei	6.03%	0.49%	Qinghai	1.68%	1.36%	Guangdong	2.01%	0.90%
7	Gansu	4.18%	0.32%	Fujian	4.04%	1.44%	Beijing	0.23%	0.10%	Sichuan	12.43%	1.15%	Chongqing	4.38%	0.87%
8	Qinghai	0.60%	0.27%	Chongqing	4.04%	0.52%	Tianjin	0.23%	0.05%	Heilongjiang	4.39%	0.85%	Zhejiang	3.65%	0.63%
9	Jilin	1.95%	0.15%	Shanghai	0.19%	0.10%	Shanghai	0.11%	0.02%	Zhejiang	3.58%	0.56%	Liaoning	1.46%	0.40%
10	Liaoning	1.13%	0.08%	Qinghai	0.38%	0.06%	Qinghai	0.34%	0.02%	Jilin	2.19%	0.39%	Hainan	0.18%	0.05%

Table 2: Summary Statistics

	# of Observations	Mean	Standard Deviation	Minimum	Maximum
Own-Account	285,479	0.323	0.467	0	1
Employer	221,611	0.127	0.333	0	1
Gender	335,600	0.581	0.493	0	1
Education	335,600	0.294	0.456	0	1
Marriage	335,600	0.788	0.409	0	1
Children	335,600	0.424	0.494	0	1
Age	335,600	33.95	9.219	15	60
Years of Stay	275,603	4.667	4.794	0	50
Hukou	334,107	0.857	0.350	0	1
Ethnity	329,139	0.966	0.182	0	1
Hang-out Dummy	129,666	0.597	0.491	0	1

Table 3A: Basic Regressions for Own-Account Self-Employed Migrants

	Dependent variable: *Ownaccount$_i$*			
	General Network		Same-Industry Network	
	SAR Linear	SAR Linear	SAR Linear	SAR Linear
	(1)	(2)	(3)	(4)
Network Own-Account	0.344***	0.115***	0.317***	0.055***
	(0.047)	(0.011)	(0.032)	(0.009)
Gender	0.033***	0.061***	0.042***	0.061***
	(0.006)	(0.004)	(0.005)	(0.004)
Education	-0.0611***	-0.060***	-0.055***	-0.060***
	(0.008)	(0.004)	(0.008)	(0.004)
Marital Status	0.104***	0.108***	0.102***	0.109***
	(0.009)	(0.006)	(0.009)	(0.006)
Having Kids	0.121***	0.070***	0.112***	0.071***
	(0.007)	(0.004)	(0.006)	(0.004)
Age	0.012***	0.010***	0.013***	0.0102***
	(0.002)	(0.002)	(0.002)	(0.002)
Age Square	-0.000***	-0.000***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
Stay Years	0.013***	0.008***	0.011***	0.009***
	(0.001)	(0.001)	(0.001)	(0.001)
Stay Years Square	-0.000***	-0.000***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
Hukou	0.062***	0.046***	0.054***	0.046***
	(0.010)	(0.005)	(0.009)	(0.005)
Ethnity	0.057*	0.011	0.057*	0.012
	(0.032)	(0.010)	(0.033)	(0.010)
Home Province FE	No	Yes	No	Yes
Host City FE	No	Yes	No	Yes
Industry FE	No	Yes	No	Yes
Year FE	No	Yes	No	Yes
# of Observations	230,014	230,013	230,014	230,013
R-Squared	0.196	0.078	0.265	0.081

Note: Standard errors are in parenthesis. ***, **, * indicate significance at the 1%, 5% and 10% level.

Table 3B: Basic Regressions for Employer Migrants

	Dependent variable: $Employer_i$			
	General Network		Same-Industry Network	
	SAR Linear	SAR Linear	SAR Linear	SAR Linear
	(1)	(2)	(3)	(4)
Network Employer	0.190**	0.106	0.106*	0.0235
	(0.092)	(0.102)	(0.064)	(0.0165)
Gender	0.025***	0.047***	0.027***	0.0475***
	(0.003)	(0.003)	(0.003)	(0.0028)
Education	0.004	-0.008***	0.005	-0.0077***
	(0.004)	(0.002)	(0.005)	(0.0019)
Marital Status	0.040***	0.055***	0.039***	0.0549***
	(0.006)	(0.004)	(0.006)	(0.0038)
Having Kids	0.077***	0.053***	0.078***	0.0535***
	(0.008)	(0.004)	(0.008)	(0.0040)
Age	0.009***	0.008***	0.009***	0.0080***
	(0.001)	(0.001)	(0.001)	(0.0011)
Age Square	-0.000***	-0.000***	-0.000***	-0.0001***
	(0.000)	(0.000)	(0.000)	(0.0000)
Stay Years	0.005***	0.004***	0.005***	0.0041***
	(0.001)	(0.001)	(0.001)	(0.0007)
Stay Years Square	-0.000	-0.000	-0.000	-0.0000
	(0.000)	(0.000)	(0.000)	(0.0000)
Hukou	-0.004	-0.003	-0.006	-0.0031
	(0.008)	(0.004)	(0.007)	(0.0041)
Ethnity	0.037**	0.007	0.040**	0.0075
	(0.015)	(0.005)	(0.017)	(0.0051)
Home Province FE	No	Yes	No	Yes
Host City FE	No	Yes	No	Yes
Industry FE	No	Yes	No	Yes
Year FE	No	Yes	No	Yes
# of Observations	174,893	174,887	174,893	174,887
R-Squared	0.095	0.050	0.095	0.0491

Note: Standard errors are in parenthesis. ***, **, * indicate significance at the 1%, 5% and 10% level.

Table 4: Own-Account Self-Employment Regressions in Four Industries

	Dependent variable: *Ownaccount$_i$*							
	Manufacturing		Wholesale and Retail		Lodging and Catering		Construction	
	General Network	Same-Industry Network	General Network	Same-Industry Network	General Network	Same-Industry Network	General Network	Same-Industry Network
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Network	0.019	0.157***	0.067***	0.045***	0.034	0.089***	0.117***	0.218***
Own-Account	(0.042)	(0.038)	(0.011)	(0.008)	(0.025)	(0.015)	(0.032)	(0.028)
Gender	0.010***	0.009***	0.094***	0.094***	0.128***	0.127***	-0.022*	-0.0203*
	(0.002)	(0.002)	(0.005)	(0.005)	(0.009)	(0.009)	(0.011)	(0.0109)
Education	-0.014***	-0.0131***	-0.070***	-0.070***	-0.066***	-0.065***	-0.038***	-0.035**
	(0.002)	(0.002)	(0.009)	(0.009)	(0.008)	(0.008)	(0.013)	(0.013)
Marital Status	0.012***	0.012***	0.231***	0.231***	0.250***	0.249***	0.067***	0.065***
	(0.002)	(0.002)	(0.017)	(0.017)	(0.012)	(0.012)	(0.014)	(0.013)
Having Kids	0.037***	0.036***	0.046***	0.046***	0.080***	0.079***	0.092***	0.090***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.010)	(0.010)	(0.010)	(0.010)
Age	0.001	0.001	0.022***	0.022***	0.023***	0.023***	0.002	0.003
	(0.001)	(0.001)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.003)
Age Square	-0.000	-0.000	-0.000***	-0.000***	-0.000***	-0.000***	-0.000	-0.000*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Stay Years	0.001	0.001	0.009***	0.009***	0.007***	0.007***	0.015***	0.014***
	(0.001)	(0.001)	(0.001)	(0.0010)	(0.002)	(0.002)	(0.002)	(0.002)
Stay Years Square	0.000	0.000	-0.000***	-0.000***	-0.000**	-0.000**	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Hukou	0.007	0.007*	0.061***	0.060***	0.034***	0.034***	0.050***	0.046***
	(0.004)	(0.004)	(0.007)	(0.007)	(0.012)	(0.012)	(0.011)	(0.012)
Ethnity	0.008*	0.007*	-0.002	-0.004	-0.067**	-0.065**	0.049***	0.047***
	(0.004)	(0.004)	(0.013)	(0.014)	(0.031)	(0.031)	(0.014)	(0.013)
Home Province FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Host City FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
# of Observations	63,393	63,393	44,514	44,514	26,947	26,947	23,525	23,525
R-Squared	0.017	0.028	0.141	0.141	0.173	0.176	0.046	0.058

Note: Standard errors are in parenthesis. ***, **, * indicate significance at the 1%, 5% and 10% level.

Table 5: Own-Account Self-Employment Regressions - Channel Examination

	Dependent variable: $Ownaccount_i$			
	Hang-out with Network People		Years of Residence in Host City	
	General Network	Same-Industry Network	General Network	Same-Industry Network
	(1)	(2)	(3)	(4)
Network Own-Account	0.086***	0.033***	0.060***	0.001
	(0.014)	(0.011)	(0.015)	(0.009)
Network*Hang-out HP	0.020*	0.027***		
	(0.010)	(0.007)		
Hang-out HP	-0.002	-0.003		
	(0.005)	(0.004)		
Network*Stay Years			0.022***	0.023***
			(0.003)	(0.003)
Network*Stay Years Square			-0.001***	-0.001***
			(0.0001)	(0.000)
Stay Years	0.008***	0.008***	0.001	0.002
	(0.001)	(0.001)	(0.0011)	(0.001)
Stay Years Square	-0.000***	-0.000***	0.0001	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Gender	0.053***	0.053***	0.061***	0.061***
	(0.005)	(0.004)	(0.004)	(0.004)
Education	-0.0520***	-0.052***	-0.059***	-0.059***
	(0.0051)	(0.005)	(0.004)	(0.004)
Marital Status	0.121***	0.122***	0.110***	0.111***
	(0.008)	(0.008)	(0.006)	(0.006)
Having Kids	0.069***	0.070***	0.0708***	0.071***
	(0.004)	(0.004)	(0.0035)	(0.003)
Age	0.011***	0.011***	0.0106***	0.011***
	(0.002)	(0.002)	(0.002)	(0.002)
Age Square	-0.000***	-0.000***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
Hukou	0.040***	0.040***	0.045***	0.045***
	(0.006)	(0.006)	(0.005)	(0.005)
Ethnity	0.014	0.015	0.012	0.012
	(0.011)	(0.012)	(0.010)	(0.010)
Home Province FE	Yes	Yes	Yes	Yes
Host City FE	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
# of Observations	63,511	63,511	230,013	230,013
R-Squared	0.071	0.074	0.080	0.085

Note: Standard errors are in parenthesis. ***, **, * indicate significance at the 1%, 5% and 10% level.

Table 6: Own-Account Self-Employment Regressions in Subsamples

	Dependent variable: $Ownaccount_i$								
	Gender		Education		Hukou		City Size		
	Male	Female	Low	High	Rural	Urban	BSGS	Large	Small
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Network Own-Account	0.333***	0.241***	0.304***	0.265***	0.306***	0.232***	0.543***	0.327***	0.274***
	(0.020)	(0.017)	(0.019)	(0.021)	(0.017)	(0.027)	(0.045)	(0.024)	(0.021)
Other Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Home Province FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Host City FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
# of Observations	138,845	91,163	164,995	65,012	199,737	30,261	42,730	98,552	88,727
R-Squared	0.106	0.092	0.087	0.089	0.098	0.085	0.100	0.104	0.101

Note: Standard errors are in parenthesis. ***, **, * indicate significance at the 1%, 5% and 10% level.

References

[1] Andersson, L., & Hammarstedt, M. (2012). Ethnic enclaves, networks and self-employment among Middle Eastern immigrants in Sweden. International Migration, 53(6), 27–40.

[2] Baltagi, B. H., & Yen, Y. F. (2014). Hospital treatment rates and spillover effects: Does ownership matter?. Regional Science and Urban Economics, 49, 193-202.

[3] Bertrand, M., Luttmer, E. F., & Mullainathan, S. (2000). Network effects and welfare cultures. The Quarterly Journal of Economics, 115(3), 1019-1055.

[4] Borjas, G. J. (1986). The self-employment experience of immigrants. The Journal of Human Resources, 21(4), 485–506.

[5] Bosma, N., Van Praag, M., Thurik, R., & De Wit, G. (2004). The value of human and social capital investments for the business performance of startups. Small Business Economics, 23(3), 227-236.

[6] Cai, F., Park A., & Zhao Y. (2008). The Chinese labor market in the reform era. Chapter6 in Laren Brandt and Thomas Rawski (eds.), China's Great Economic Transformation, Cambridge University Press, 167-214.

[7] Calvo, G. A., & Wellisz, S. (1980). Technology, entrepreneurs, and firm size. The Quarterly Journal of Economics, 95(4), 663-677.

[8] Carr, D. (1996). Two paths to self-employment? Women's and men's self-employment in the United States, 1980. Work and occupations, 23(1), pp.26-53.

[9] Chen, ., Jin G. Z., & Yue Y. (2010). Peer migration in China. NBER Working Paper #15671 .

[10] Clark, K., Drinkwater, S., & Robinson, C. (2017). Self-employment amongst migrant groups: new evidence from England and Wales. Small Business Economics, 48(4), 1047–1069.

[11] Constant, A., & Zimmermann, K. F. (2006). The making of entrepreneurs in Germany: Are native men and immigrants alike?. Small business economics, 26(3), 279-300.

[12] Cueto, B., & Rodríguez Álvarez, V. (2015). Determinants of immigrant self-employment in Spain. International Journal of Manpower, 36(6), 895–911.

[13] Dawson, C., Henley, A., & Latreille, P. (2014). Individual motives for choosing self-employment in the UK: Does region matter?. Regional studies, 48(5), 804-822.

[14] Fairlie, R. W. & Lofstrom, M. (2015). Immigration and entrepreneurship. In Handbook of the economics of international migration (Vol. 1, pp. 877-911). North-Holland.

[15] Gagnon, J.J., Xenogiani, T. & Xing, C. (2012). Are all migrants really worse off in urban labour markets? New empirical evidence from China.

[16] Gindling, T. H., & Newhouse, D. (2014). Self-employment in the developing world. World Development, 56, 313–331.

[17] Gurley-Calvez, T., Biehl, A., & Harper, K. (2009) Time-use patterns and women entrepreneurs. American Economic Review 99: 139-144.

[18] Hammarstedt, M., & Shukur, G. (2009). Testing the home-country self-employment hypothesis on immigrants in Sweden. Applied Economics Letters 16: 745-748.

[19] Kelejian, H. H., & Prucha I. R. (1998). A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. The Journal of Real Estate Finance and Economics 17.1: 99-121.

[20] Kelejian, H. H., Prucha, I. R., & Yuzefovich Y. (2004). Instrumental variable estimation of a spatial autoregressive model with autoregressive disturbances: Large and small sample results. Advances in Econometrics: Spatial and Spatio-Temporal Econometrics: 163-198.

[21] Kerr, W. R., & M. Mandorff (2015). Social networks, ethnicity, and entrepreneurship. NBER Working Paper 21597.

[22] Klier, T., & McMillen, D. P. (2008). Clustering of auto supplier plants in the United States. Journal of Business & Economic Statistics 26.4: 460–471

[23] Lee, L. (2003). Best spatial two-stage least squares estimators for a spatial autoregressive model with autoregressive disturbances. Econometric Reviews 22.4: 307-335.

[24] LeSage, J., & Pace, R. K. (2009). Introduction to spatial econometrics. Chapman and Hall/CRC Press.

[25] Lin, X. (2010). Identifying peer effects in student academic achievement by spatial autoregressive models with group unobservables. Journal of Labor Economics 28.4: 825-860.

[26] Lin, J., G. Wang, & Y. Zhao (2004). Regional inequality and labor transfers in China. Economic Development and Cultural Changes, 52(3): 587-603.McKenzie, David and Hillel Rapoport. 2007. Network effects and the dynamics of migration and inequality: Theory and evidence from Mexico. Journal of Development Economics, 84(1): 1-24.

[27] Martín-Montaner, J., Serrano-Domingo, G., & Requena-Silvente, F. (2018). Networks and self-employed migrants. Small Business Economics, 51(3), 735-755.

[28] Mata, F. & Pendakur, R. (1999). Immigration, labor force integration and the pursuit of self-employment. International migration review, 33(2), pp.378-402.

[29] Meng, X., & Zhang, J. (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.

[30] Park, A. (2008). Rural-urban inequality in China. China Urbanizes: Consequences, Strategies, and Policies, 41-63.

[31] Raijman, R., & Tienda, M. (2000). Immigrants' pathways to business ownership: A comparative ethnic perspective. International migration review, 34(3), pp.682-706.

[32] Simoes, N., Crespo, N., & Moreira, S. B. (2016). Individual determinants of self-employment entry: what do we really know? Journal of Economic Surveys, 30(4), 783–806.

[33] Van der Sluis, J., Van Praag, C. M., & Vijverberg, W. (2008). Education and entrepreneurship selection and performance: A review of the empirical literature. Journal of economic surveys, 22(5), 795-841.

[34] Van Praag, C. M., & Van Ophem, H. (1995). Determinants of willingness and opportunity to start as an entrepreneur. Kyklos, 48(4), 513-540.

[35] Wellington, A.J. (2006). Self-employment: the new solution for balancing family and career?. Labour Economics, 13(3), pp.357-386.

[36] Yuengert, A. (1995). Testing hypotheses of immigrant self-employment. Journal of Human Resources 30: 194-204.

[37] Zhang, J., & Zhao, Z. (2015). Social-family network and self-employment: evidence from temporary rural–urban migrants in China. IZA Journal of Labor & Development, 4(1), pp.1-21.

[38] Zhao, Y. (2003). The role of migrant networks in labor migration: The case of China. Contemporary Economic Policy, 21(4), 500-511.

[39] Zhao, Z. (2004). Rural-urban Migration in China-What do we know and what do we need to know? CHINA ECONOMIC QUARTERLY-BEIJING, 3, pp.517-536.

[40] Zhu, N. (2002). The impacts of income gaps on migration decisions in China. China Economic Review, 13(2-3): 213-230.