

Network Effects on Migrants' Labor Market Outcomes in China

Shan Jin*

First version April 26, 2020; this version October 29, 2020

Abstract

This paper uses 2005 China's mini-census to study how networks affect migrants' migration decisions, employment, and wage levels. Different from existing studies, this paper takes into account the existence of self-selection in the labor market. With the help of theoretical model, we have a better understanding of the mechanism of networks as well as the different network effects on rural and urban migrants. We find out that networks affect both rural and urban migrants' migration decisions positively. In terms of employment, networks exert positive impacts on rural migrants but not on urban ones, which is due to the different quality drops between rural and urban migrants when the networks increase. Such employment effects also lead rural migrants to face a more severe negative wage impact than urban migrants.

Keywords: migrants, network, wage, China

JEL Classification Numbers: J31, J61, O15, O18.

*Department of Economics, 3117 Pamplin Hall, 880 West Campus Drive Virginia Tech, Blacksburg, VA, 24061. Email: jshan3@vt.edu.

1 Introduction

Labor migration starts rising since the late 1980s in China. Nowadays, with the perfection of migration supporting system, more and more people try to get job opportunities outside their hometowns. For such people, network in the destination labor market exerts an indispensable role. Networks not only affect migrants' migration decisions, but also influence various labor market outcomes in unspectacular ways. Figure 1 shows us the percentages of move-in and move-out in each province, which indicates an unbalanced migration situation in China. Migrants in popular destinations (the eastern coast region) or from central regions are more likely to have greater networks in the destinations when migrating. Some people say that a large network size in the destination helps them settle down more easily, while others are concern about a high competition along with large networks in destinations. Thus, how to choose a proper destination is worth discussing.

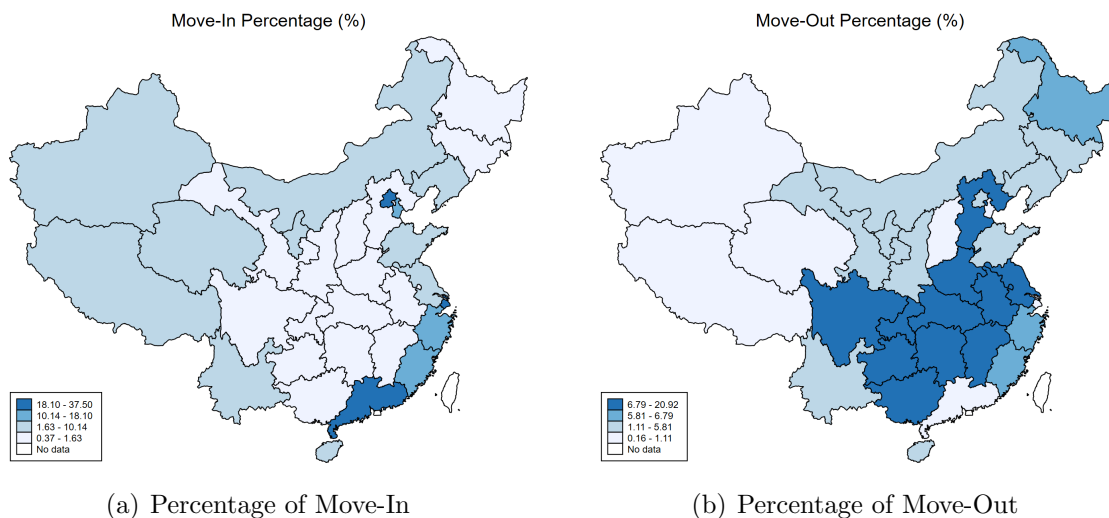


Figure 1: Migration Status (Data from 2005 1% Population Survey)

Studies about how networks affect labor market outcomes are many and varied, such as network effects on migrants' migration decisions (Patel and Vella, 2013; Kerr and Mandorff, 2015), network effects on migrants' employment status (Munshi, 2003; Giulietti et al., 2018), and some studies even discuss wage impacts (Patel and Vella, 2013; Zhang et al., 2012; Long et al., 2017; Chen et al., 2018). Obviously, networks play a role that cannot be ignored in the labor market. Most of the existing studies focus on a single aspect of labor market outcomes, which may present only a one-sided picture of the network effects. If we put the pieces together, we would have a more

complete story, and have a better understanding of network effects on migrants. So, in this paper, we try to not only find out how networks affect migration decisions, but also examine the employment as well as the wage impacts together. Especially, in terms of wage impact, we pay attention to a general network impact on all migrants instead of the network impact on network users. Different from existing studies, this paper takes into account the existence of self-selection in the labor market. As a result, a change in network size no longer indicates an increase or decrease of identical migrants.

Except for this, another point of this paper is that we explore the difference of network effects between rural and urban migrants. As we know, when we mention migrants, spotlights are only on the rural migrants. However, as the migrants with urban *hukou* status keep rising from 1990, there has already been a sizeable urban migrants group in the labor market. Thus, we try to fill the gap, and find out how networks give rural and urban migrants different impacts on the various labor market outcomes.

First of all, we build up a theoretical foundation in this paper. Based on Beaman (2012) and Calvo-Armengol and Jackson (2004)'s framework, we study the network effects on migrants' employment and wage levels theoretically. Taking the case when there is a positive selection, and when there is a positive relationship between networks and the percentage of employed early migrants as an example¹, we conclude that employment effects of networks depends on the type of network indirect impact, and also depends on the comparison between networks' direct and indirect impact on employment². If there is a negative network indirect impact, employment rate decreases as networks increase. If there is a positive network indirect impact, then employment rate decreases when the size of direct impact exceeds indirect impact. In terms of network effects on wage, if the early employed migrants' wage levels are higher than the average workers' wage levels in the market, networks expose a positive wage impact.

In the following part of this paper, we use 2005 China's mini-census to explore network effects on labor market outcomes in China. First of all, we check the type of self-selection as well as the relationship between networks and the percentage of employed early migrants, so that corresponding theoretical propositions could help us in explaining the empirical findings. We then turn to the network effects on various labor market outcomes. From the regression results, we realize that networks affect migration decisions for both rural and urban migrants positively. What's more, networks positively affect overall employment. Nevertheless, when we look over the difference

¹Other cases are also discussed in the theory model section.

²Direct impact is how network affects the probability of being employed directly, while indirect impact represents the network impact on the probability of being employed through social network.

between rural and urban migrants, the results indicate that networks only help rural migrants to be employed, but have negative effect on urban ones. With the help of the theory model, we attribute such difference to the different quality drops of rural and urban migrants when networks increase. In terms of the wage effects of networks, generally speaking, networks affect migrants' wage levels negatively. Although both rural and urban migrants face negative network wage effects, the magnitudes are different. Rural migrants have a more severe negative effects than urban ones, which is because there are more unqualified rural workers who are able to get jobs via networks when networks increase. As a result, negative wage effect for urban migrants mainly reflects the decreasing quality of urban migrants (due to the positive selection), but the wage decreases for rural migrants has already exceeded quality drops from selection.

In this paper, we fill the gap of existing literature to some extent. As we mentioned, most studies ignore the existence of self-selection in the labor market, and focus on a single aspect of labor market outcomes, which is very likely to read the results incompletely. Also, if we do not consider urban migrants in the study, we are not able to understand migrants group very well. With the help of theoretical model, we try to study a complete mechanism behind the networks in the labor market.

In the next section, we present some related literature about network effects on various labor market outcomes. Section 3 shows the theoretical foundation. Section 4 introduces the data and the sample. Section 5 gives us empirical evidence and according explanations. The last section concludes.

2 Related Literature

There are numbers of factors prompt people to migrate, such as higher wages, more amenities, or better education quality for next generations in destinations. Besides such factors, one's networks in the destination could also be a crucial one to affect their migration decisions. Some people think a higher network in the destination is helpful, since employed fellows can provide additional job opportunities and help new migrants settle down as soon as possible. Such people would like to migrate to the place where there is a large size of network. However, others take a higher network for a fierce competition. People from the same origin may have similar geographical characteristics, so migrants from the same origin would be more likely to be good at or to choose to work in the same industry or occupations (Patel and Vella, 2013; Hashimzade et al., 2014; Kerr and Mandorff, 2015). For such individuals, they prefer to migrate to a destination with fewer origin fellows. It seems like people's opposite

choices on migration destinations are due to their different expectations on networks. Some people believe that networks have positive effects on their labor market outcomes, while others believe not.

According to the existing literature, effects of network on labor market outcomes are varied. Dolfin and Genicot (2010) explore the mechanism of how networks affect migration decisions, and point out networks help migrants in many aspects. The most crucial point to promote people to migrate is that larger networks provide more job information. Similarly, in China's labor market, researchers also find out that as networks increase, people are more likely to migrate (Zhao, 2003; Du et al., 2005; Chen et al., 2010). Migrants acquire more information on migration experience, on job searching, or even on having financial support by networks. After all, networks are used to reduce migrants' migration costs (McKenzie and Rapoport, 2007) when people are making migration decisions.

In terms of network effects in destination labor markets, one of the findings is that networks offer help for raising employment rate of migrants. Munshi (2003) explores the network effects on Mexican migrants' employment status, and points out that networks positively affect migrants in the U.S. labor market. *Guanxi*, a Chinese word which indicates networks, increases the probability of being employed in non-farm positions for migrants as well (Zhang and Li, 2003). Calvo-Armengol and Jackson (2004); Calvo-Armengol and Jackson (2007); Ioannides and Soetevent (2006) provide some theoretical support on such positive network effects. Moreover, there is also a longer tenure for the migrants who get jobs via networks (Loury, 2006). However, when we focus on different types of networks, there may have some other conclusions. For example, Beaman (2012) mentions that only tenured network members increase migrants' employment, but the new arrivals cannot. In addition, Giulietti et al. (2018) suggest what help migrants in acquiring jobs are strong ties which play different roles from weak ties among migrants.

There exists some evidence to say people get more job information via networks, and the crucial reason why the job information passed from early migrants is valuable for new entrants is because migrants from the same origin are more likely to choose similar jobs (Patel and Vella, 2013). Attitudes and beliefs, such as risk preference, honesty weight, could affect people's occupational choices (Hashimzade et al., 2014), and migrants who come from the same origin also show similar sociodemographic characteristics. As a result, network referral effects seems to be higher among migrants from the same origin (Bayer et al., 2008), and the effects are more obvious in self-employment (Allen, 2000; Zhang and Zhao, 2015). Kerr and Mandorff (2015) also suggest that a smaller group has a higher entrepreneurial concentration, which once again confirms the referral effects among migrants from the same origin.

Network effects are not only found in migrants' employment, but there are also some intuitive effects on migrants' wage levels. Yogo (2011) and Giulietti et al. (2010) both point out that higher networks increase migrants wage levels if the migrants use networks to get jobs. What's more, Patel and Vella (2013) also conclude that if new arrivals choose the most popular jobs for their according early migrants, there exists a wage premium for new entrants. Such network effects could also explain a part of wage inequality between ethnicity (Arrow and Borzekowski, 2004).

However, conclusions of network effects on wages are not consistent. For example, Beaman (2012) suggests that positive network wage effects only come from early migrants. In the meantime, Ye et al. (2012) focus only on high level networks, and find out that the higher level networks increase migrants' wage in the labor market. Even beyond that, there are also couple of researches put forward some other opinions on network wage effects. Zhang et al. (2012) realizes that using network cannot increase all migrants' wage level, Long et al. (2017) and Chen et al. (2018) even find out a negative network effect on wages in informal job searching.

As mentioned above, effects of networks on labor market outcomes are varied. Migrants may benefit from a higher network, but may also be adversely affected. Also, when we focus on different migrants groups, there could exist various network effects. How networks affect rural and urban migrants differently is also worth discussing. It is undeniable that relationship plays an important role in the labor market. Networks affect people's employment, people's wage level, or even the population distribution structure of a society. Therefore, a correct understanding of network effects on the labor market is valuable and necessary for all of us.

3 The Model

Based on Beaman (2012) and Calvo-Armengol and Jackson (2004)'s framework, we build a model to discuss the network effects on migrants' labor market outcomes.

3.1 Network Effects on Migrants' Employment

Assume that, at the beginning of this period, there are $(1 - b)$ of the early migrants have already been employed, where $b \in [0, 1]$. Thus, there are three kinds of migrants in the labor market - employed early migrants, unemployed early migrants and new arrivals. The unemployed early and new migrants can get job offers from employers directly (direct way). In addition, since there exists employed early migrants, the

unemployed early migrants and new migrants are able to have job offers passed from unemployed early migrants (indirect way).

We use net_c to represent the size of network for migrants from community c in the destination labor market, where $c \in C$. According to Munshi (2003), networks are defined as the proportion of the community located at destination, and $net_c \in [0, 1]$. Each community c has a total population of N_c , so $(net_c \times N_c)$ represents the population of migrants from community c in the destination.

Let $e(net_c)$ denote the percentage of early migrants, so that $(e(net_c) \times net_c \times N_c)$ is the number of early migrants. We use en_c to represent the network that comes from early migrants, while $rn_c(en_c)$ is the network from recent migrants, and the size of recent network is correlated with the size of early network. Therefore, we have

$$\begin{aligned} net_c &= en_c + rn_c(en_c) \\ &= e(net_c) \times net_c + rn_c(en_c), \end{aligned}$$

where $rn'_c(en_c) \geq 0$ ³. To explore how the percentage of early migrants $e(net_c)$ changes with net_c , we take the first derivative of $e(net_c)$, and find out

$$\begin{aligned} \frac{\partial e(net_c)}{\partial net_c} &= \frac{\partial \frac{en_c}{net_c}}{\partial net_c} \\ &= \frac{1}{net_c^2} \times \frac{rn_c(en_c) - rn'_c(en_c)en_c}{1 + rn'_c(en_c)}. \end{aligned}$$

Thus, when $\frac{rn'_c(en_c)}{rn_c(en_c)} \leq \frac{1}{en_c}$, there is a positive relationship between networks and the percentage of early migrants ($e'(net_c) \geq 0$), otherwise we have a negative relationship⁴.

s_c denotes the probability of being employed for unemployed migrants from community c ,

$$s_c = a(net_c) + r_c, \tag{1}$$

where $a(net_c)$ is the probability of being employed in a direct way, while r_c represents the probability of being employed through an indirect way, $a(net_c) \in [0, 1]$, $r_c \in [0, 1]$. We assume that employed early migrants pass the offer she receives via networks. As long as individual i is unemployed, i accept the offer that passed from an employed early migrant. How does network affect $a(net_c)$ depends on the average ability change

³According to Zhao (2003); Du et al. (2005); Chen et al. (2010), a high existing network in the destination prompts people to migrate.

⁴See Appendix A.

with the network. For example, if there is a positive self-selection on migrants, the average ability of migrants from community c decreases as the network size increases. As a result, the probability of receiving job offers from employers decreases when net_c increases. Therefore, we have $a'(net_c) \leq 0$, $a''(net_c) \geq 0$. If there is a negative self-selection, we have $a'(net_c) \geq 0$, $a''(net_c) \leq 0$. For no selection case, we have $a'(net_c) = 0$.

The probability of being employed through an indirect way (r_c) can be written as

$$r_c = \frac{(1-b)e(net_c)a(net_c)}{1-(1-b)e(net_c)}.$$

Thus, the probability of being employed is

$$\begin{aligned} s_c &= a(net_c) + r_c, \\ &= \frac{a(net_c)}{1-(1-b)e(net_c)}. \end{aligned}$$

As can be seen from above, migrants have a direct impact ($a(net_c)$), and an indirect impact (r_c) from networks when applying for jobs. Thus, how networks affect migrants' employment depends on the size of direct and indirect impact.

3.1.1 Case 1: when there is no self-selection

In this case, we know that the probability of being employed in a direct way does not change as the network size increases ($a'(net_c) = 0$). This is also the case which most existing studies discuss.

Proposition 1. *If $e'(net_c) \geq 0$, s_c (weakly) increases as net_c increases, $\forall net_c \in [0, 1]$.
If $e'(net_c) < 0$, s_c (weakly) decreases as net_c increases, $\forall net_c \in [0, 1]$.*

Proof. See appendix. □

3.1.2 Case 2: when there is a positive self-selection

If there is a positive self-selection on migrants, we have $a'(net_c) \leq 0$.

Proposition 2. *If $e'(net_c) \geq 0$, s_c (weakly) increases as net_c increases when $\frac{a'(net_c)}{a(net_c)} \in [-\frac{(1-b)e'(net_c)}{1-(1-b)e(net_c)}, 0]$; s_c (weakly) decreases as net_c increases when $\frac{a'(net_c)}{a(net_c)} \in (-\infty, -\frac{(1-b)e'(net_c)}{1-(1-b)e(net_c)}]$.
If $e'(net_c) < 0$, s_c (weakly) decreases as net_c increases, $\forall net_c \in [0, 1]$.*

Proof. See appendix. □

3.1.3 Case 3: when there is a negative self-selection

If there is a negative self-selection on migrants in the destination labor market, we have $a'(net_c) \geq 0$.

Proposition 3. *If $e'(net_c) \geq 0$, s_c (weakly) increases as net_c increases, $\forall net_c \in [0, 1]$.*

If $e'(net_c) < 0$, s_c (weakly) increases as net_c increases when $\frac{a'(net_c)}{a(net_c)} \in [-\frac{(1-b)e'(net_c)}{1-(1-b)e(net_c)}, +\infty)$; s_c (weakly) decreases as net_c increases when $\frac{a'(net_c)}{a(net_c)} \in [0, -\frac{(1-b)e'(net_c)}{1-(1-b)e(net_c)}]$.

Proof. See appendix. □

For a better understanding, we take the case when $e'(net_c) \geq 0$ and when there is a positive self-selection on migrants as an example. As Table 1 shows, in the destination labor market, migrants face both direct and indirect impacts from networks. We know that as the network increases, there is a negative direct impact ($a'(net_c) \leq 0$). If as network increases, indirect impact decreases, we find a negative network impact on migrants' employment. However, if as network increases, indirect impact increases, then we have to discuss. When direct impact dominates, migrants have a negative network impact, but when indirect impact dominates, migrants have a positive network impact on employment.

Table 1: Different Impacts on Migrants

| Conditions | Direct & Indirect Impact | Overall Impact |
|--|--|-------------------------|
| Network \uparrow Indirect Impact \downarrow | Negative Direct Impact Negative Indirect Impact | Negative Network Impact |
| Network \uparrow Indirect Impact \uparrow Direct Impact \geq Indirect Impact | Negative Direct Impact Positive Indirect Impact | Negative Network Impact |
| Network \uparrow Indirect Impact \uparrow Direct Impact $<$ Indirect Impact | Negative Direct Impact Positive Indirect Impact | Positive Network Impact |

3.2 Network Effects on Migrants' Wage

In terms of jobs with different wage levels, individual i can receive a job offer with wage w_i in the destination labor market. We assume that if i is unemployed, she takes any offer she gets. If i has already been employed at the beginning of the period with wage w_i^* , she compares the wage she receives, w_i , with her own wage w_i^* . If $w_i \geq w_i^*$, she takes this offer, otherwise, she passes the offer to others in her network. Employed early migrant i earns the same wage as at the beginning of the period if she does not take any other offer.

Wage has an *i.i.d.* distribution from $[w_{low}, w_{high}]$ with a cumulative distribution function $F(x)$. $G(x)$ is a cumulative distribution function of employed early migrants' wage.

3.2.1 Case 1: when there is no self-selection

In this case, the probability of being employed in a direct way does not change with the network size, so $a'(net_c) = 0$.

Proposition 4. *If $e'(net_c) \geq 0$, all employed migrants' average wage (weakly) increases as net_c increases when G first-order stochastically dominates F ($G(x) \leq F(x)$); all employed migrants' average wage (weakly) decreases as net_c increases when $G(x) \geq F(x)$.*

If $e'(net_c) < 0$, all employed migrants' average wage (weakly) increases as net_c increases when $G(x) \geq F(x)$; all employed migrants' average wage (weakly) decreases as net_c increases when $G(x) \leq F(x)$.

Proof. See appendix. □

3.2.2 Case 2: when there is a positive self-selection

In this case, we know that $a'(net_c) \leq 0$.

Proposition 5. *If $e'(net_c) \geq 0$, all employed migrants' average wage (weakly) increases as net_c increases when $G(x) \leq F(x)$; all employed migrants' average wage (weakly) decreases as net_c increases when $G(x) \geq F(x)$.*

If $e'(net_c) < 0$, all employed migrants' average wage (weakly) increases as net_c increases when $\frac{a'(net_c)}{a(net_c)} \geq \frac{e'(net_c)}{e(net_c)}$ and $G(x) \geq F(x)$, or when $\frac{a'(net_c)}{a(net_c)} \leq \frac{e'(net_c)}{e(net_c)}$ and $G(x) \leq F(x)$; all employed migrants' average wage (weakly) decreases as net_c increases when $\frac{a'(net_c)}{a(net_c)} \geq \frac{e'(net_c)}{e(net_c)}$ and $G(x) \leq F(x)$, or when $\frac{a'(net_c)}{a(net_c)} \leq \frac{e'(net_c)}{e(net_c)}$ and $G(x) \geq F(x)$.

Proof. See appendix. □

Therefore, taking the case when there is a positive self-selection and $e'(net_c) \geq 0$ as an example, if the quality of early employed migrants is higher than the average quality in the market, networks expose positive wage effects; otherwise, migrants' wage decreases as networks increase.

3.2.3 Case 3: when there is a negative self-selection

In this case, there is a positive relationship between network size and the probability of being employed in a direct way. Therefore, $a'(net_c) \geq 0$.

Proposition 6. *If $e'(net_c) \geq 0$, all employed migrants' average wage (weakly) increases as net_c increases when $\frac{a'(net_c)}{a(net_c)} \geq \frac{e'(net_c)}{e(net_c)}$ and $G(x) \geq F(x)$, or when $\frac{a'(net_c)}{a(net_c)} \leq \frac{e'(net_c)}{e(net_c)}$ and $G(x) \leq F(x)$; all employed migrants' average wage (weakly) decreases as net_c increases when $\frac{a'(net_c)}{a(net_c)} \geq \frac{e'(net_c)}{e(net_c)}$ and $G(x) \leq F(x)$, or when $\frac{a'(net_c)}{a(net_c)} \leq \frac{e'(net_c)}{e(net_c)}$ and $G(x) \geq F(x)$.*

If $e'(net_c) < 0$, all employed migrants' average wage (weakly) increases as net_c increases when $G(x) \geq F(x)$; all employed migrants' average wage (weakly) decreases as net_c increases when $G(x) \leq F(x)$.

Proof. See appendix. □

From the model, we realize that how network affects migrants employment and wage depends on the relationship between networks and the percentage of early migrants ($e'(net_c)$), as well as the type of self-selection ($a'(net_c)$) in the labor market.

4 Data and Descriptive Statistics

In this paper, we use the China's 2005 mini-census data which is published by the National Bureau of Statistics of China. As a national survey data, the survey covers 1% of the national population, and the data set has various information for each subject, such as subjects' gender, age, education status, marital status, household registration information, employment status, and even has wage information, which help us broaden our research.

Since we have to acquire subjects' networks from household registration information and residence information, subjects who do not have household registration or residence information are not included. According to one's household registration information (*hukou*), subjects can be classified into rural *hukou* or urban *hukou* type. However, besides the *hukou* type, there is also a *hukou* location which indicates one's place of residence (origin). For most of the people in China, it is difficult to change the *hukou* location. Therefore, in a specific place, one can also be categorized into a local or a migrant according to the *hukou* location.

We mainly focus on the working age population whose ages are between 16 and 60 in this study. Subjects who are not able to work, self report to be retired or in school are dropped from the data set. Also, people who report to be employed but

the income resources are unemployment insurance or property income are also not considered as our objects. Migrants are defined by their household registration and residence information. A subject is a migrant if she works in a different province from their household registration place (province). There are two reasons for why we focus on inter-provincial migrants. First of all, inter-provincial migration is mainstream in the labor market of China. As Hou et al. (2005) shows, the vast majority of migrants choose the east region as their migration destinations in China. Li et al. (2014) also conclude that there is an uptrend in inter-provincial migration in China since 1985. The amount of inter-provincial migrants in 2000-2005 is 38,042,340, which is almost four times of the number in ten years ago. Another reason is that most people are used to distinguish themselves by province in social settings. Thus, inter-provincial migration is more proper in network analysing.

According to the education status, we are able to acquire subjects' education years, and categorize the sample into three educational levels: compulsory and below education level (middle and primary school or below), high school level, college and above level. What's more, we have three sectors in the data set which are state-owned enterprises, private enterprises, and other ownership. Based on previous research, industries are classified into primary industry, manufacturing industry, construction industry, basic service industry, and advanced service industry (Ge and Yang, 2009). Hourly wage is acquired by monthly wage and working hours from the data set.

Table 2 shows the basic characteristics of the sample. From Table 2, we find out that migrants are much younger than locals, and there are significant fewer married subjects among migrants. When we separate the migrants into urban and rural sample, we notice that urban migrants even have more education years than locals. Employed rates are slightly higher for migrants, since migrants are more eager to make money to survive in destinations.

In terms of the employed sample in our data set, migrants' mean working hours are greater than that of locals. For rural migrants, the mean hourly wage is not higher than urban locals' which indicates rural migrants are more likely to work in low-paid positions. Self-employment rate is similar for both locals and migrants, and rural migrants have a slightly higher self-employment rate. From the table, we also notice that urban locals are mainly working in state-owned enterprises, while migrants are more likely to choose to work in private enterprises. Besides sectors, we also have an industry distribution in Table 2. The most popular industries for migrants are basic service industry and manufacturing industry, while urban locals are employed most in advanced service industry.

Table 2: Characteristics of Locals and Migrants

| | Local | | Migrant | |
|---------------------------|---------|-----------|---------|--------|
| | Urban | Rural | Urban | Rural |
| <i>All</i> | | | | |
| Male | 0.53 | 0.49 | 0.52 | 0.52 |
| Age | 37.96 | 38.16 | 31.85 | 30.04 |
| Married | 0.83 | 0.82 | 0.63 | 0.65 |
| Years of Education | 11.47 | 7.19 | 12.18 | 8.52 |
| Employed | 0.76 | 0.89 | 0.85 | 0.89 |
| Number of Respondent | 386,746 | 1,017,144 | 18,233 | 91,184 |
| <i>Employed</i> | | | | |
| Weekly Working Hours | 45.88 | 45.70 | 48.58 | 55.18 |
| Hourly Wage | 6.82 | 2.45 | 10.61 | 4.71 |
| Self-Employed | 0.12 | 0.09 | 0.10 | 0.13 |
| <i>Ownership</i> | | | | |
| State-Owned Enterprise | 0.58 | 0.03 | 0.20 | 0.09 |
| Private Enterprise | 0.33 | 0.90 | 0.59 | 0.67 |
| Other Ownership | 0.09 | 0.06 | 0.20 | 0.24 |
| <i>Industry</i> | | | | |
| Primary Industry | 0.06 | 0.76 | 0.01 | 0.04 |
| Manufacturing Industry | 0.25 | 0.10 | 0.34 | 0.57 |
| Construction Industry | 0.04 | 0.03 | 0.06 | 0.09 |
| Advanced Service Industry | 0.36 | 0.02 | 0.21 | 0.05 |
| Basic Service Industry | 0.29 | 0.09 | 0.38 | 0.26 |
| Number of Respondent | 292,455 | 905,957 | 15,525 | 81,317 |

Data comes from 2005 1% national population survey

From the above summary statistics, we learn about our sample distribution, and also have some idea about the subjects' labor market performance. It is undeniable that networks play an important role in our culture as well as in the society. Thus, how migrants' labor market outcomes are affected by networks is worth a special study and discussion. In the next section, we provide some empirical evidence to show how the networks affect various labor market outcomes, and also present different network effects on rural and urban migrants.

5 Empirical Implications

In this section, we explore the network effects on migrants' migration decisions, employment status, and wages. Before turning to the network effects on the various labor market outcomes, we have to check several theoretical conditions, such as the type of self-selection, and the relationship between networks and the percentage of employed early migrants ($e'(net_c)$).

5.1 Type of Selection in the Labor Market

First of all, we propose to explore the type of self-selection on migrants in the labor market. We use the Mincerian equation (Mincer, 1958) as the basic structure. The specification can be shown as

$$\ln W_i = \beta_0 + \mathbf{X}_i\delta + \beta_1 MIG_i + \varepsilon_i, \quad (2)$$

where W_i represents i 's hourly wage, \mathbf{X}_i is a vector which indicates one's gender, working experience, marital status, ethnicity, educational status, industries and so on, MIG_i is a dummy variable which indicates i 's migration status, and ε_i is an error term with zero mean.

Table 3 shows the basic OLS regression results. In the regression, subjects are separated into three educational level - compulsory and below level, high school level, college and above level. The compulsory and below level includes the subjects who have no educational experience or attend up to middle school; high school level stands for the subjects with high school educational experience; college and above level represents the subjects with more than high school education (both college and graduates school). We also categorize industries into 5 groups (Ge and Yang, 2009) - primary industry, manufacturing industry, construction industry, advanced service industry, and basic service industry; and categorize sectors into 3 groups - state-owned enterprises, private enterprises, and other ownership. For eliminating the effect from unequal educational resources, we use the pupil-teacher ratios in primary school level, middle school level, high school level, and educational funding data of migrants' origin communities to control for education quality in each province.⁵ Destination fixed effects are also controlled in the regression.

⁵Similar to Card and Krueger (1992), we use pupil-teacher ratios and education funding as proxies for education quality. Data comes from China Statistical Yearbooks.

Table 3: Self-Selection on Rural & Urban Migrants

| | (1) All Sample | (2) Rural Sample | (3) Urban Sample |
|-------------------------------------|---------------------|---------------------|---------------------|
| Male | 0.203*** (0.001) | 0.207*** (0.001) | 0.186*** (0.002) |
| Working Experience | 0.009*** (0.000) | 0.005*** (0.000) | 0.019*** (0.000) |
| Married | 0.073*** (0.002) | 0.092*** (0.002) | 0.029*** (0.003) |
| Migrant | 0.085*** (0.004) | 0.105*** (0.004) | 0.265*** (0.007) |
| Urban <i>Hukou</i> | 0.105*** (0.002) | | |
| Migrant \times Urban <i>Hukou</i> | 0.214*** (0.006) | | |
| High School | 0.173*** (0.002) | 0.128*** (0.002) | 0.260*** (0.003) |
| College | 0.640*** (0.003) | 0.538*** (0.010) | 0.707*** (0.003) |
| Province Dummies | Yes | Yes | Yes |
| Constant | 0.740*** (0.015) | 0.626*** (0.017) | 1.223*** (0.039) |
| Observations | 1295254 | 987274 | 307980 |
| r2_a | 0.522 | 0.391 | 0.411 |

Standard errors in parentheses

Working experience square, ethnicity, industry, ownership type, and education qualities are also controlled

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

From Table 3 we know that male and married status both provide a wage premium. Also, as can be seen from the table, wage becomes higher for the one with a higher education level. In the first column of Table 3, subjects with high school educational experience earn 0.173 more than those with lower education level, and people who have been in college earn almost 0.467 more than high school graduates. Additionally, from the regression results we learn that, the one who works in the advanced service industry is more likely to earn the highest wage, and state-owned enterprises are better-paid than other sectors.

One of the most important findings in Table 3 is that both rural and urban migrants earn more than their according locals when we control for other related factors. Consistent with Ge and Jin (2020), we are able to say there is a positive selection on both rural and urban migrants in the labor market. This indicates that the average quality of migrants decreases as the proportion of the community located at destination (network) increases.

In Table 4, we show an evidence of another theoretical condition that we have in the propositions of section 3. Relationship between the percentage of (employed) early migrants and networks are shown in the following table. Based on the model settings, we

define networks as the proportion of the community located at destination⁶. Moreover, we have different employed early migrants in the two columns based on the migration years⁷. As can be seen from the table, both give us significant positive correlations.

Table 4: Relationship between Networks and the Proportion of Employed Early Migrants

| | (1) Proportion of Employed Early Migrants (Define More than 4 Years as Early Migrants) | (2) Proportion of Employed Early Migrants (Define More than 3 Years as Early Migrants) |
|------------------|---|---|
| Network | 0.231*** (0.011) | 0.266*** (0.010) |
| Province Dummies | Yes | Yes |
| Constant | 0.338*** (0.001) | 0.414*** (0.001) |
| Observations | 32263 | 41182 |
| r2.a | 0.345 | 0.285 |

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

So far we have the empirical evidence on the type of self-selection as well as the relationship between networks and the percentage of employed early migrants. So we currently expect that corresponding theoretical propositions would help us in explaining the empirical findings in the next section. In the following section, we find out how networks affect migrants' various labor market outcomes, and also pay attention to the difference of network effects between rural and urban migrants in the labor market of China.

5.2 Network Effects on Labor Market Outcome

5.2.1 Network Effect on Migration Decisions

Networks could affect numbers of decisions made from migrants. One of the most basic decisions for migrants is whether to become migrants or not. Therefore, in Table 5, we first explore the network effect on migrants' migration decisions.

We use the specification as

$$NEWMIG_{icm} = \beta_0 + \beta_1 EARLYNET_{icm} + \delta PRO_{im} + \varepsilon_i, \quad (3)$$

⁶According to Munshi (2003), network is calculated by the number of migrants from origin community c (province level) who work in destination m (province level) divided by the population of origin community c .

⁷According to Zhao (2003), the one who has migrated more than 48 months can be regarded as "early migrants." For robustness checking, we also regress 3+ year migrants on networks in column (2).

where $NEWMIG_{icm}$ is normalized new migrant i who come from origin community c work in destination m , $EARLYNET_{icm}$ indicates i 's early networks which is the proportion of employed early migrants within an origin community c located at destination m , PRO_{im} is a destination province indicator, and ε_i is an error term with zero mean.

Table 5: Network Effect on New Migrants' Migration Decisions

| | Normalized New Migrants (≤ 4 Years) | | | Normalized New Migrants (≤ 3 Years) | | |
|---|---|----------------------|----------------------|---|----------------------|----------------------|
| | All (1) | Rural (2) | Urban (3) | All (4) | Rural (5) | Urban (6) |
| Early Network (from 4+ Years Migrants) | 57.454*** (0.132) | 56.983*** (0.145) | 59.659*** (0.350) | | | |
| Early Network (from 3+ Years Migrants) | | | | 44.850*** (0.111) | 44.499*** (0.122) | 46.562*** (0.297) |
| Province Dummies | Yes | Yes | Yes | Yes | Yes | Yes |
| Constant | -0.987*** (0.001) | -0.984*** (0.001) | -0.996*** (0.002) | -0.978*** (0.001) | -0.974*** (0.001) | -0.990*** (0.002) |
| Observations | 71608 | 61134 | 10474 | 61687 | 52915 | 8772 |
| r2_a | 0.904 | 0.904 | 0.902 | 0.908 | 0.907 | 0.910 |

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 5 shows the according regression results. Column (1) to (3) is the regression results of a case when we define new migrants as the ones who migrate to current destination within 4 years, while the following three columns is for new migrants who migrate no more than 3 years. Column (2) and (5) are for rural sample, while (3) and (6) are for urban sample only. As can be seen from the table, numbers of new migrants increase when the early networks raise. In other words, if there exists a high proportion of the community located at destination, that promotes the migrants' fellows to migrate. This result is true for both rural and urban migrants.

5.2.2 Network Effect on Employment

Apart from migration decisions, we are also interest in the network effect on migrants' employment. We use the data in hand to find out whether networks help migrants to be employed more or not. Moreover, we are able to explore whether the network effect is different for rural and urban migrants. Table 6 and 7 show us empirical evidence to answer these questions.

In Table 6, we first do a probit regression to explore how networks affect migrants' employment status. The specification can be written as

$$EMPLOYED_i = \beta_0 + \mathbf{X}_i\delta + \beta_1NET_i + \varepsilon_i, \quad (4)$$

where $EMPLOYED_i$ indicates whether a migrant is employed or not, \mathbf{X}_i is the same

vector as we use in equation (2), NET_i represents migrants' network size, and ε_i is an error term with zero mean.

Table 6: Network Effect on Migrants' Employment Status (Probit)

| | All Migrants | | Recent Migrants | |
|------------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| | General (1) | Rural/Urban (2) | General (3) | Rural/Urban (4) |
| Network | 0.380** (0.062)** | 0.609*** (0.054)* | 0.504** (0.076)** | 0.683*** (0.070)** |
| Urban <i>Hukou</i> × Network | | -1.374*** (-0.245)*** | | -1.202*** (-0.195)* |
| Urban <i>Hukou</i> | -0.236*** (-0.038)*** | -0.182*** (-0.043)*** | -0.201*** (-0.030)*** | -0.153*** (-0.034)*** |
| High School | 0.048*** (0.008)*** | 0.049*** (0.008)*** | 0.027 (0.004) | 0.027 (0.004) |
| College | 0.176*** (0.027)*** | 0.170*** (0.026)*** | 0.066* (0.010)* | 0.059 (0.009) |
| Male | 1.038*** (0.168)*** | 1.038*** (0.168)*** | 0.950*** (0.144)*** | 0.950*** (0.144)*** |
| Age | 0.091*** (0.015)*** | 0.090*** (0.015)*** | 0.103*** (0.016)*** | 0.103*** (0.016)*** |
| Married | -0.647*** (-0.105)*** | -0.646*** (-0.105)*** | -0.748*** (-0.113)*** | -0.747*** (-0.113)*** |
| Province Dummies | Yes | Yes | Yes | Yes |
| Constant | -0.802*** | -0.832*** | -1.001*** | -1.022*** |
| Observations | 109,417 | 109,417 | 72,053 | 72,053 |

Margins in parentheses

age square, ethnicity, and education qualities are also controlled

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 6 shows us the regression results of equation (4). We use all migrants as our sample in the first two columns, and then restrict the sample to recent migrants⁸ in column (3) and (4). As the table shows, networks help all migrants generally in being employed. No matter using all migrants or recent migrants as our sample, we are able to find out that networks are helpful to rural migrants, but cannot help urban migrants to be employed. Among recent migrants, the marginal effect of network for rural migrants is 0.070, while that for urban migrants is -0.125. Additionally, urban workers are less likely to be employed when comparing with rural ones. Perhaps rural migrants have higher moving cost, so they are more eager to make money in the destination. In terms of education attainment level, the more educated, the more likely to be employed. Moreover, male or single migrants are relatively favorable in the labor market.

From the above regression, we find out networks help rural migrants but cannot help urban migrants in being employed. Let us recall the theoretical settings in section

⁸We define migrants who migrate no more than 4 years as recent migrants in Table 6.

3. In equation (1), we assume that there is only one kind of migrants in the market. However, in reality, we are able to categorize migrants into rural and urban ones. Therefore, for migrants in group k , the probability of being employed can be expressed as:

$$s_{kc} = a_k(net_c) + \frac{(1-b)e(net_c)a(net_c)}{1-(1-b)e(net_c)},$$

where $k \in \{rural, urban\}$. Network indirect impact will be the same for both rural and urban migrants. Since migrants could have job offers passed from both rural and urban employed early migrants. Thus, we suppose that, for current rural and urban job seekers (migrants), the difference in network effect is mainly due to the difference in direct impact ($a'_{rural}(net_c)$ vs. $a'_{urban}(net_c)$). In other words, different network employment effect for current rural and urban job seekers (migrants) comes from the different quality drops for the two groups of migrants.

Different from the probability of being employed for job seekers, let w_c denote the probability of being employed for all the migrants in the labor market:

$$\begin{aligned} w_c &= (1-b)e(net_c) + [1-(1-b)e(net_c)]s_c, \\ &= (1-b)e(net_c) + a(net_c). \end{aligned}$$

Similar to the s_{kc} ,

$$w'_{kc} = (1-b)e'(net_c) + a'_k(net_c),$$

where $k \in \{rural, urban\}$. $e'(net_c)$ is the same for both rural and urban migrants. Thus, the different network employment effect for all rural and urban migrants also comes from the different quality drops for the two groups of migrants ($a'_{rural}(net_c)$ vs. $a'_{urban}(net_c)$).

Table 7 not only supports the existence of positive selection in the labor market, but also shows the different quality drops for rural and urban migrants as networks increase. We regress overall, rural, urban migrants' average education years on networks, and the results indicate that when the network size increases, migrants' average education years go down significantly. No matter for the rural or the urban migrants, migrants' average ability decreases as the migration rate increases, and the average quality drop for urban migrants is greater than rural migrants ($a'_{urban}(net_c) < a'_{rural}(net_c) < 0$.)

Table 7: Relationship between Networks and Average Education Years for Migrants

| | (1) Average Education Years All | (2) Average Education Years Rural | (3) Average Education Years Urban |
|------------------|---------------------------------------|---|---|
| Network | -8.885*** (0.076) | -3.163*** (0.054) | -10.765*** (0.206) |
| Province Dummies | Yes | Yes | Yes |
| Constant | 10.423*** (0.009) | 9.082*** (0.007) | 13.508*** (0.020) |
| Observations | 109417 | 91184 | 18233 |
| r2_a | 0.402 | 0.498 | 0.409 |

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Since overall network employment effect is positive (Table 6), we conclude that, for current job seekers (migrants), indirect impact is positive (because direct impact $a'(net_c) \leq 0$). When the size of indirect impact is the same for both rural and urban migrants, the one with smaller direct impact has a smaller (even negative) network effect on employment. What's more, for all of the rural and urban migrants in the labor market, we know that $(1 - b)e'(net_c) \geq 0$. Thus, the one with smaller $a'_k(net_c)$ has a smaller (even negative) network effect on employment.

Up to now, we explore network effects on migration decision and employment status. Besides this, how networks affect migrants' wage remains a mystery. In the following subsection, we will show empirical evidence on the network effects on migrants' wage, as well as how rural and urban migrants affected differently by networks.

5.2.3 General Network Effect on Wage

In this section, we discuss the network effect on migrants' log hourly wage. From the following regressions, we learn about whether migrants with a higher network have a wage premium or not in the destination labor market. Based on the classic Mincerian equation (Mincer, 1958), the log wage is composed as

$$\ln W_i = \beta_0 + \mathbf{X}_i\delta + \beta_1 NET_i + \varepsilon_i, \quad (5)$$

where $\ln W_i$ is a log hourly wage of individual i , \mathbf{X}_i a vector which stands for personal characteristics, NET_i is migrants' network size, and ε_i is an error term with zero mean.

According regression results are shown in Table 8. In the first column of Table 8, we do the basic OLS regression. Nevertheless, there still exists some factors which affect both network size and wage levels. For dealing with the endogeneity problem, we adopt a lagged network size as an instrument variable (IV), and the 2SLS regression results are

shown in the second column of Table 8. The instrument variable is from China’s 2000 Census data, and the data coverage is similar to the 2005 mini-census. Additionally, the lagged network size is correlated with the current migration proportion, but cannot affect other aspects, such as economic conditions, in 2005 ⁹.

Table 8: Network Effect on Migrants’ Wage

| | OLS Results (1) | 2SLS Results (2) |
|--------------------|----------------------|----------------------|
| Network | -0.865*** (0.051) | -0.819*** (0.053) |
| Male | 0.155*** (0.003) | 0.155*** (0.003) |
| Working Experience | 0.015*** (0.001) | 0.015*** (0.001) |
| Married | 0.039*** (0.005) | 0.039*** (0.005) |
| Urban <i>Hukou</i> | 0.217*** (0.006) | 0.218*** (0.006) |
| High School | 0.259*** (0.005) | 0.260*** (0.005) |
| College | 0.879*** (0.010) | 0.881*** (0.008) |
| Province Dummies | Yes | Yes |
| Constant | 0.874*** (0.027) | 0.869*** (0.025) |
| Observations | 96842 | 96275 |
| r _{2.a} | 0.362 | 0.361 |

Standard errors in parentheses

Working experience square, ethnicity, industry, ownership type, and education qualities are also controlled

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

From Table 8, we are able to realize that networks decrease migrants’ wage levels. Even after we do the 2SLS regression, the negative coefficient remains similarly. As we know that there is a positive selection on migrants in the labor market¹⁰, which means as networks increase, the average quality of migrants falls. As a result, when networks increase, migrants’ wage drops by 0.819. Besides the network impact, we find out that, among migrants, male, married or urban *Hukou* holders are more likely to have a relatively higher wage. Also, migrants with more working experience or higher educational attainment level are also better paid. As shown in Table 8, migrants with high school education earn 0.260 more than those with compulsory education or lower, and college graduates have 0.621 wage premium when comparing to high school

⁹Hausman test result strongly rejects the null hypothesis. The first-stage results of 2SLS regression indicate that there is a strong correlation between network and its instrument variable. Cragg-Donald Wald F statistic shows there is no weak instrument problem.

¹⁰From Table 3 and 7.

graduates. From the regression results we are also able to conclude that the best-paid industry is the advanced service industry, while the worst-paid one is, as we expected, the primary industry. In terms of sectors, the largest wage gap exists between state-owned enterprises and private enterprises.

So far, we notice that there is a negative network effect on migrants' wage. Then let us recall what we have in the theoretical model. In Proposition 5, we mention that all employed migrants' average wage (weakly) decreases as network increases when: $G(x) \geq F(x)$ ¹¹. Therefore, in Figure 2(a) and 2(b), we check whether we have such condition in our data¹².

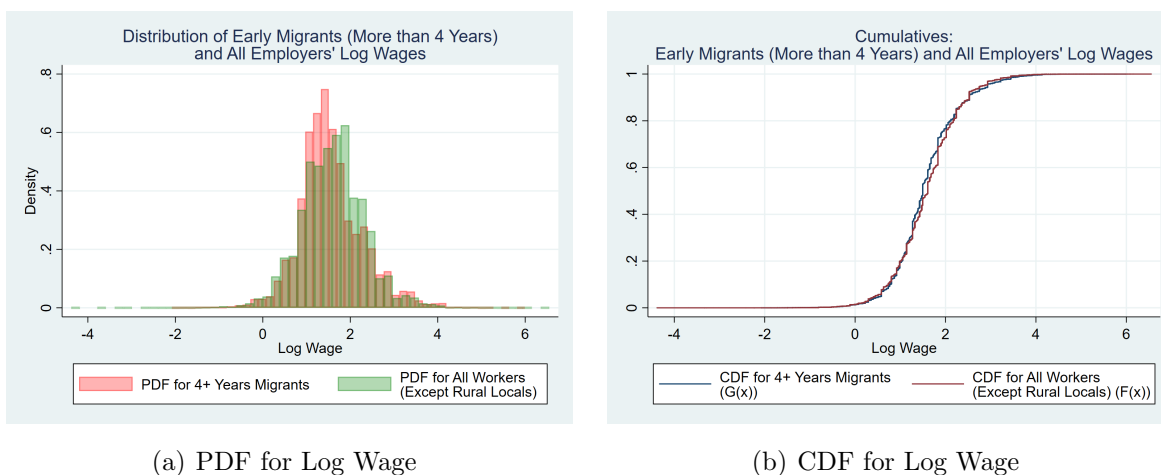


Figure 2: Distributions for Early Migrants and All Employers' Log Wage

Figure 2 gives us a PDF graph and an according CDF graph for log wage. In both figures, we treat people who migrate to current destination more than 4 years as early migrants. Figure 2(a) shows that early migrants' wage distributions skew leftward when comparing to the wage distribution in the labor market. Accordingly, we also find out a slightly leftward moved CDF curves for employed early migrants¹³. Therefore, after having a negative network wage effect empirically, we observe its theoretical condition from the data as well.

¹¹Wage has an *i.i.d.* distribution from $[w_{low}, w_{high}]$ with a cumulative distribution function $F(x)$. $G(x)$ is a cumulative distribution function of employed early migrants' wage.

¹²The reason why we exclude rural locals in both figures is that the vast majority of migrants migrate out for non-farming jobs, but from the data we realize that more than 70% of local rural workers live by farming. Therefore, we suppose that migrants' main competitors are urban locals.

¹³Kolmogorov–Smirnov tests statistic for panel (b) is 0.0687 [0.000].

5.2.4 Network Effect on Wage for Rural and Urban Migrants

From the previous subsection, we realize that networks have a negative effect on migrants' wage. In this subsection, we pay further attention to whether there is any difference between rural and urban migrants, and try to figure out the reason why migrants are facing different network effects.

Table 9: Network Effect on Wage for Rural and Urban Migrants

| | (1) | (2) |
|------------------------------|----------------------|----------------------|
| | OLS Results | 2SLS Results |
| Network | -0.935*** (0.051) | -0.900*** (0.055) |
| Urban <i>Hukou</i> × Network | 0.489*** (0.112) | 0.544*** (0.101) |
| Urban <i>Hukou</i> | 0.195*** (0.008) | 0.193*** (0.007) |
| Male | 0.155*** (0.003) | 0.155*** (0.003) |
| Working Experience | 0.015*** (0.001) | 0.015*** (0.001) |
| Married | 0.039*** (0.005) | 0.038*** (0.005) |
| High School | 0.260*** (0.005) | 0.260*** (0.005) |
| College | 0.881*** (0.010) | 0.883*** (0.008) |
| Province Dummies | Yes | Yes |
| Constant | 0.885*** (0.027) | 0.882*** (0.025) |
| Observations | 96842 | 96275 |
| r2_a | 0.362 | 0.361 |

Standard errors in parentheses
Working experience square, ethnicity, industry,
ownership type, and education qualities are
also controlled
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

We adopt an interaction term of *urban* × *network* in Table 9. Similar to the previous table, we do a simple OLS regression in column (1), and do a 2SLS regression in the second column. From the table, we notice that networks still have negative effects on both rural and urban migrants. What's more, as we would expect, networks affect rural and urban migrants differently. When networks increase, rural migrants' wage drops by 0.900, while that of urban migrants only decreases by 0.356. The results indicate that networks have a more severe impact on rural migrants. From section 5.2.2 we know that networks help rural migrants to be employed. As a result, as networks increase, more rural recent migrants who were not able to get jobs by themselves now are working. Therefore, as networks increase, the wage decrease for urban migrants

mainly reflects the decreasing quality of urban migrants¹⁴, but the wage decreases for rural migrants exceed the quality drops from positive selection. That is because there are more unqualified rural migrants get jobs via networks as the networks increase.

6 Conclusion

By using 2005 China's mini-census, this paper studies how networks affect migrants' migration decisions, employment, as well as wage levels. Moreover, this paper also pays attention to the existence of self-selection and the difference between rural and urban migrants, so that we consider the heterogeneity of migrants, and have a better understanding of the roles that networks play in the labor market of China.

The empirical findings indicate that networks give both rural and urban migrants a positive impact on their migration decisions. For employment, networks only exert positive impacts on rural migrants, but not on urban ones. As networks increase, urban migrants even face a lower employment rate. With the help of theoretical framework, we are able to find out the reason for such rural-urban difference. This is due to the different quality drops between rural and urban migrants when the networks increase. Such employment effect has also led to different negative network wage effects for rural and urban migrants. Since, among rural migrants, there are more unqualified workers get jobs via networks, the wage decreases for rural migrants has already exceeded quality drops from positive selection.

¹⁴Due to the positive selection.

A Appendix A

How does the percentage of early migrants $e(net_c)$ change with net_c :

$$\begin{aligned}
\frac{\partial e(net_c)}{\partial net_c} &= \frac{\partial \frac{en_c}{net_c}}{\partial net_c} = \frac{\frac{\partial en_c}{\partial net_c} net_c - en_c}{net_c^2} \\
&= \frac{1}{net_c^2} \left[\frac{\partial [net_c - rn_c(en_c)]}{\partial net_c} net_c - en_c \right] \\
&= \frac{1}{net_c^2} \left[net_c - \frac{\partial rn_c(en_c)}{\partial net_c} net_c - net_c + rn_c(en_c) \right] \\
&= \frac{1}{net_c^2} \left[rn_c(en_c) - \frac{\partial rn_c(en_c)}{\partial [en_c + rn_c(en_c)]} net_c \right].
\end{aligned}$$

Let $en_c + rn_c(en_c) = k(en_c) = y$, $en_c = k^{-1}(y)$. Therefore,

$$\begin{aligned}
\frac{\partial rn_c(en_c)}{\partial [en_c + rn_c(en_c)]} &= \frac{\partial [y - k^{-1}(y)]}{\partial y} = 1 - \frac{1}{k'[k^{-1}(y)]} \\
&= 1 - \frac{1}{k'(en_c)} = 1 - \frac{1}{1 + rn'_c(en_c)}.
\end{aligned}$$

$$\begin{aligned}
\frac{\partial e(net_c)}{\partial net_c} &= \frac{1}{net_c^2} \left[rn_c(en_c) - \left(1 - \frac{1}{1 + rn'_c(en_c)} \right) \times net_c \right] \\
&= \frac{1}{net_c^2} \times \frac{rn_c(en_c) - rn'_c(en_c)en_c}{1 + rn'_c(en_c)}.
\end{aligned}$$

B Appendix B

B.1 Case 1: when there is no self-selection

Proposition 1. *If $e'(net_c) \geq 0$, s_c (weakly) increases as net_c increases, $\forall net_c \in [0, 1]$.*

If $e'(net_c) < 0$, s_c (weakly) decreases as net_c increases, $\forall net_c \in [0, 1]$.

Proof.

$$\frac{\partial s_c}{\partial net_c} = \frac{a'(net_c)[1 - (1 - b)e(net_c)] + (1 - b)e'(net_c)a(net_c)}{[1 - (1 - b)e(net_c)]^2}, \quad (6)$$

$$= \frac{a'(net_c) - (1 - b)[e(net_c)a'(net_c) - e'(net_c)a(net_c)]}{[1 - (1 - b)e(net_c)]^2}. \quad (7)$$

We know that there is no self-selection on migrants ($a'(net_c) = 0$).

(1) If $e'(net_c) \geq 0$, we have $a'(net_c) - (1-b)[e(net_c)a'(net_c) - e'(net_c)a(net_c)] = (1-b)e'(net_c)a(net_c) \geq 0$, $\frac{\partial s_c}{\partial net_c} \geq 0$.

(2) If $e'(net_c) < 0$, $a'(net_c) - (1-b)[e(net_c)a'(net_c) - e'(net_c)a(net_c)] = (1-b)e'(net_c)a(net_c) \leq 0$, $\frac{\partial s_c}{\partial net_c} \leq 0$.

□

B.2 Case 2: when there is a positive self-selection

Proposition 2. *If $e'(net_c) \geq 0$, s_c (weakly) increases as net_c increases when $\frac{a'(net_c)}{a(net_c)} \in [-\frac{(1-b)e'(net_c)}{1-(1-b)e(net_c)}, 0]$; s_c (weakly) decreases as net_c increases when $\frac{a'(net_c)}{a(net_c)} \in (-\infty, -\frac{(1-b)e'(net_c)}{1-(1-b)e(net_c)})$.
If $e'(net_c) < 0$, s_c (weakly) decreases as net_c increases, $\forall net_c \in [0, 1]$.*

Proof. In this case, the probability of being employed in a direct way decreases as the network increases, so $a'(net_c) \leq 0$.

(1) According to equation (8), if $e'(net_c) \geq 0$:

when $a'(net_c) - (1-b)[e(net_c)a'(net_c) - e'(net_c)a(net_c)] \geq 0 \Leftrightarrow a'(net_c) \geq -\frac{(1-b)e'(net_c)a(net_c)}{1-(1-b)e(net_c)}$, $\frac{\partial s_c}{\partial net_c} \geq 0$. Therefore, we have $\frac{\partial s_c}{\partial net_c} \geq 0$ when $\frac{a'(net_c)}{a(net_c)} \in [-\frac{(1-b)e'(net_c)}{1-(1-b)e(net_c)}, 0]$;

when $a'(net_c) - (1-b)[e(net_c)a'(net_c) - e'(net_c)a(net_c)] \leq 0 \Leftrightarrow a'(net_c) \leq -\frac{(1-b)e'(net_c)a(net_c)}{1-(1-b)e(net_c)}$, $\frac{\partial s_c}{\partial net_c} \leq 0$. Therefore, we have $\frac{\partial s_c}{\partial net_c} \leq 0$ when $\frac{a'(net_c)}{a(net_c)} \in (-\infty, -\frac{(1-b)e'(net_c)}{1-(1-b)e(net_c)})$.

(2) If $e'(net_c) < 0$, $a'(net_c) - (1-b)[e(net_c)a'(net_c) - e'(net_c)a(net_c)] \leq 0$, $\frac{\partial s_c}{\partial net_c} \leq 0$.

□

B.3 Case 3: when there is a negative self-selection

Proposition 3. *If $e'(net_c) \geq 0$, s_c (weakly) increases as net_c increases, $\forall net_c \in [0, 1]$.*

If $e'(net_c) < 0$, s_c (weakly) increases as net_c increases when $\frac{a'(net_c)}{a(net_c)} \in [-\frac{(1-b)e'(net_c)}{1-(1-b)e(net_c)}, +\infty)$; s_c (weakly) decreases as net_c increases when $\frac{a'(net_c)}{a(net_c)} \in [0, -\frac{(1-b)e'(net_c)}{1-(1-b)e(net_c)}]$.

Proof. In this case, we have $a'(net_c) \geq 0$.

(1) If $e'(net_c) \geq 0$, we have $a'(net_c) - (1-b)[e(net_c)a'(net_c) - e'(net_c)a(net_c)] \geq 0$, $\frac{\partial s_c}{\partial net_c} \geq 0$.

(2) According to equation (8), if $e'(net_c) < 0$:

when $a'(net_c) - (1-b)[e(net_c)a'(net_c) - e'(net_c)a(net_c)] \geq 0 \Leftrightarrow a'(net_c) \geq -\frac{(1-b)e'(net_c)a(net_c)}{1-(1-b)e(net_c)}$, $\frac{\partial s_c}{\partial net_c} \geq 0$. Since $a'(net_c) \geq 0$, we have $\frac{\partial s_c}{\partial net_c} \geq 0$ when $\frac{a'(net_c)}{a(net_c)} \in [-\frac{(1-b)e'(net_c)}{1-(1-b)e(net_c)}, +\infty)$;

when $a'(net_c) - (1-b)[e(net_c)a'(net_c) - e'(net_c)a(net_c)] \leq 0 \Leftrightarrow a'(net_c) \leq -\frac{(1-b)e'(net_c)a(net_c)}{1-(1-b)e(net_c)}$, $\frac{\partial s_c}{\partial net_c} \leq 0$. Therefore, we have $\frac{\partial s_c}{\partial net_c} \leq 0$ when $\frac{a'(net_c)}{a(net_c)} \in [0, -\frac{(1-b)e'(net_c)}{1-(1-b)e(net_c)}]$.

□

B.4 Case 1: when there is no self-selection

Proposition 4. *If $e'(net_c) \geq 0$, all employed migrants' average wage (weakly) increases as net_c increases when G first-order stochastically dominates F ($G(x) \leq F(x)$); all employed migrants' average wage (weakly) decreases as net_c increases when $G(x) \geq F(x)$.*

If $e'(net_c) < 0$, all employed migrants' average wage (weakly) increases as net_c increases when $G(x) \geq F(x)$; all employed migrants' average wage (weakly) decreases as net_c increases when $G(x) \leq F(x)$.

Proof. Let us discuss the proportion of employed individuals whose wages are greater than x for all job seekers, $\forall x$. For all job seekers,

$$\begin{aligned}
\% \text{ of } \frac{\text{wage} > x}{\text{employed}} &= \frac{\% \text{ early employed wage} > x}{\% \text{ early employed} + (\text{early unemployed} + \text{new}) \times \% \text{ of being employed}} \\
&+ \frac{[\text{early employed wage} < x + (\text{early unemployed} + \text{new})] \times \% \text{ of current wage} > x}{\% \text{ early employed} + (\text{early unemployed} + \text{new}) \times \% \text{ of being employed}}, \\
&= \frac{(1-b)e(net_c)(1-G(x))}{(1-b)e(net_c) + [1 - (1-b)e(net_c)] \times (a(net_c) + r_c)} \\
&+ \frac{[1 - (1-b)e(net_c)(1-G(x))] \times [a(net_c)(1-F(x)) + \frac{a(net_c)(1-b)e(net_c)(1-G(x))}{1-(1-b)e(net_c)(1-G(x))}(1-F(x))]}{(1-b)e(net_c) + [1 - (1-b)e(net_c)] \times (a(net_c) + r_c)}, \\
&= \frac{(1-b)e(net_c)(1-G(x)) + a(net_c)(1-F(x))}{(1-b)e(net_c) + a(net_c)}.
\end{aligned}$$

$$\frac{\partial \% \text{ of } \frac{\text{wage} > x}{\text{employed}}}{\partial \text{net}_c} = \frac{[(1-b)(1-G(x))e'(\text{net}_c) + (1-F(x))a'(\text{net}_c)][(1-b)e(\text{net}_c) + a(\text{net}_c)]}{[(1-b)e(\text{net}_c) + a(\text{net}_c)]^2} \quad (8)$$

$$- \frac{[(1-b)e'(\text{net}_c) + a'(\text{net}_c)][(1-b)e(\text{net}_c)(1-G(x)) + a(\text{net}_c)(1-F(x))]}{[(1-b)e(\text{net}_c) + a(\text{net}_c)]^2}, \quad (9)$$

$$= \frac{(1-b)(1-F(x))[e(\text{net}_c)a'(\text{net}_c) - e'(\text{net}_c)a(\text{net}_c)]}{[(1-b)e(\text{net}_c) + a(\text{net}_c)]^2} \quad (10)$$

$$+ \frac{(1-b)(1-G(x))[a(\text{net}_c)e'(\text{net}_c) - e(\text{net}_c)a'(\text{net}_c)]}{[(1-b)e(\text{net}_c) + a(\text{net}_c)]^2}, \quad (11)$$

$$= \frac{(1-b)}{[(1-b)e(\text{net}_c) + a(\text{net}_c)]^2} [e(\text{net}_c)a'(\text{net}_c) - e'(\text{net}_c)a(\text{net}_c)][G(x) - F(x)]. \quad (12)$$

we know that, in this case, $a'(\text{net}_c) = 0$.

(1) According to equation (13), if $e'(\text{net}_c) \geq 0$,

when $G(x) - F(x) \geq 0 \Leftrightarrow G(x) \geq F(x)$, we have $\frac{\partial \% \text{ of } \frac{\text{wage} > x}{\text{employed}}}{\partial \text{net}_c} \leq 0$;

when $G(x) - F(x) \leq 0 \Leftrightarrow G(x) \leq F(x)$, we have $\frac{\partial \% \text{ of } \frac{\text{wage} > x}{\text{employed}}}{\partial \text{net}_c} \geq 0$.

(2) If $e'(\text{net}_c) < 0$,

when $G(x) - F(x) \geq 0 \Leftrightarrow G(x) \geq F(x)$, we have $\frac{\partial \% \text{ of } \frac{\text{wage} > x}{\text{employed}}}{\partial \text{net}_c} \geq 0$;

when $G(x) - F(x) \leq 0 \Leftrightarrow G(x) \leq F(x)$ or, we have $\frac{\partial \% \text{ of } \frac{\text{wage} > x}{\text{employed}}}{\partial \text{net}_c} \leq 0$.

□

B.5 Case 2: when there is a positive self-selection

Proposition 5. *If $e'(\text{net}_c) \geq 0$, all employed migrants' average wage (weakly) increases as net_c increases when $G(x) \leq F(x)$; all employed migrants' average wage (weakly) decreases as net_c increases when $G(x) \geq F(x)$.*

If $e'(\text{net}_c) < 0$, all employed migrants' average wage (weakly) increases as net_c

increases when $\frac{a'(net_c)}{a(net_c)} \geq \frac{e'(net_c)}{e(net_c)}$ and $G(x) \geq F(x)$, or when $\frac{a'(net_c)}{a(net_c)} \leq \frac{e'(net_c)}{e(net_c)}$ and $G(x) \leq F(x)$; all employed migrants' average wage (weakly) decreases as net_c increases when $\frac{a'(net_c)}{a(net_c)} \geq \frac{e'(net_c)}{e(net_c)}$ and $G(x) \leq F(x)$, or when $\frac{a'(net_c)}{a(net_c)} \leq \frac{e'(net_c)}{e(net_c)}$ and $G(x) \geq F(x)$.

Proof. In this case, $a'(net_c) \leq 0$. Based on equation (13),

(1) if $e'(net_c) \geq 0$,

when $G(x) - F(x) \geq 0 \Leftrightarrow G(x) \geq F(x)$, we have $\frac{\partial\% \text{ of } \frac{\text{wage} > x}{\text{employed}}}{\partial net_c} \leq 0$;

when $G(x) - F(x) \leq 0 \Leftrightarrow G(x) \leq F(x)$ or, we have $\frac{\partial\% \text{ of } \frac{\text{wage} > x}{\text{employed}}}{\partial net_c} \geq 0$.

(2) If $e'(net_c) < 0$,

when $\frac{a'(net_c)}{a(net_c)} \geq \frac{e'(net_c)}{e(net_c)}$ and $G(x) \geq F(x)$, or when $\frac{a'(net_c)}{a(net_c)} \leq \frac{e'(net_c)}{e(net_c)}$ and $G(x) \leq F(x)$, we have $\frac{\partial\% \text{ of } \frac{\text{wage} > x}{\text{employed}}}{\partial net_c} \geq 0$;

when $\frac{a'(net_c)}{a(net_c)} \geq \frac{e'(net_c)}{e(net_c)}$ and $G(x) \leq F(x)$, or when $\frac{a'(net_c)}{a(net_c)} \leq \frac{e'(net_c)}{e(net_c)}$ and $G(x) \geq F(x)$, we have $\frac{\partial\% \text{ of } \frac{\text{wage} > x}{\text{employed}}}{\partial net_c} \leq 0$.

□

B.6 Case 3: when there is a negative self-selection

Proposition 6. *If $e'(net_c) \geq 0$, all employed migrants' average wage (weakly) increases as net_c increases when $\frac{a'(net_c)}{a(net_c)} \geq \frac{e'(net_c)}{e(net_c)}$ and $G(x) \geq F(x)$, or when $\frac{a'(net_c)}{a(net_c)} \leq \frac{e'(net_c)}{e(net_c)}$ and $G(x) \leq F(x)$; all employed migrants' average wage (weakly) decreases as net_c increases when $\frac{a'(net_c)}{a(net_c)} \geq \frac{e'(net_c)}{e(net_c)}$ and $G(x) \leq F(x)$, or when $\frac{a'(net_c)}{a(net_c)} \leq \frac{e'(net_c)}{e(net_c)}$ and $G(x) \geq F(x)$.*

If $e'(net_c) < 0$, all employed migrants' average wage (weakly) increases as net_c increases when $G(x) \geq F(x)$; all employed migrants' average wage (weakly) decreases as net_c increases when $G(x) \leq F(x)$.

Proof. In this case, $a'(net_c) \geq 0$. Based on equation (13),

(1) if $e'(net_c) \geq 0$,

when $\frac{a'(net_c)}{a(net_c)} \geq \frac{e'(net_c)}{e(net_c)}$ and $G(x) \geq F(x)$, or when $\frac{a'(net_c)}{a(net_c)} \leq \frac{e'(net_c)}{e(net_c)}$ and $G(x) \leq F(x)$, we have $\frac{\partial\% \text{ of } \frac{\text{wage} > x}{\text{employed}}}{\partial net_c} \geq 0$;

when $\frac{a'(net_c)}{a(net_c)} \geq \frac{e'(net_c)}{e(net_c)}$ and $G(x) \leq F(x)$, or when $\frac{a'(net_c)}{a(net_c)} \leq \frac{e'(net_c)}{e(net_c)}$ and $G(x) \geq F(x)$, we have $\frac{\partial\% \text{ of } \frac{\text{wage} > x}{\text{employed}}}{\partial net_c} \leq 0$.

(2) If $e'(net_c) < 0$,

when $G(x) - F(x) \geq 0 \Leftrightarrow G(x) \geq F(x)$, we have $\frac{\partial\% \text{ of } \frac{\text{wage} > x}{\text{employed}}}{\partial net_c} \geq 0$;

when $G(x) - F(x) \leq 0 \Leftrightarrow G(x) \leq F(x)$ or, we have $\frac{\partial\% \text{ of } \frac{\text{wage} > x}{\text{employed}}}{\partial net_c} \leq 0$.

□

So, with different conditions, the proportion of employed workers whose wages are greater than x (weakly) increases or decreases with network increases for $\forall x$. Since x can be any non negative numbers, we conclude that, with different conditions, the average wage of employed workers (weakly) increases or decreases with network increases.

References

- Allen, W. D. (2000). Social Networks and Self-Employment. *The Journal of Socio-Economics*, 29(5):487–501.
- Arrow, K. J. and Borzekowski, R. (2004). Limited Network Connections and the Distribution of Wages.
- Bayer, P., Ross, S. L., and Topa, G. (2008). Place of Work and Place of Residence: Informal Hiring Networks and Labor Market Outcomes. *Journal of Political Economy*, 116(6):1150–1196.
- Beaman, L. A. (2012). Social Networks and the Dynamics of Labour Market Outcomes: Evidence from Refugees Resettled in the U.S. *The Review of Economic Studies*, 79(1):128–161.
- Calvo-Armengol, A. and Jackson, M. O. (2004). The Effects of Social Networks on Employment and Inequality. *American Economic Review*, 94(3):426–454.
- Calvo-Armengol, A. and Jackson, M. O. (2007). Networks in Labor Markets: Wage and Employment Dynamics and Inequality. *Journal of Economic Theory*, 132(1):27–46.
- Card, D. and Krueger, A. B. (1992). Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States. *Journal of Political Economy*, 100(1):1–40.
- Chen, Y., Jin, G. Z., and Yue, Y. (2010). Peer Migration in China. Technical report, National Bureau of Economic Research.
- Chen, Y., Wang, L., and Zhang, M. (2018). Informal Search, Bad Search?: The Effects of Job Search Method on Wages among Rural Migrants in Urban China. *Journal of Population Economics*, 31(3):837–876.
- Dolfin, S. and Genicot, G. (2010). What Do Networks Do? The Role of Networks on Migration and “Coyote” Use. *Review of Development Economics*, 14(2):343–359.
- Du, Y., Park, A., and Wang, S. (2005). Migration and Rural Poverty in China. *Journal of Comparative Economics*, 33(4):688–709.
- Ge, S. and Jin, S. (2020). Self-Selection and the Wage of Migrants in China. *Working Paper*.

- Ge, S. and Yang, D. T. (2009). Accounting for Rising Wages in China. *Manuscript, Virginia Tech and Chinese University of Hong Kong.*
- Giulietti, C., Guzi, M., Zhao, Z., Zimmermann, K. F., et al. (2010). Social Networks and the Labour Market Outcomes of Rural to Urban Migrants in China. *Unpublished Manuscript.*
- Giulietti, C., Wahba, J., and Zenou, Y. (2018). Strong versus Weak Ties in Migration. *European Economic Review*, 104:111–137.
- Hashimzade, N., Myles, G. D., Page, F., and Rablen, M. D. (2014). Social Networks and Occupational Choice: The Endogenous Formation of Attitudes and Beliefs about Tax Compliance. *Journal of Economic Psychology*, 40:134–146.
- Hou, W., Bao, S., and Shi, A. (2005). Analysis of the Changing Spatial Patterns of Migration in China. *Population Science of China*, 5.
- Ioannides, Y. M. and Soetevent, A. R. (2006). Wages and Employment in A Random Social Network with Arbitrary Degree Distribution. *American Economic Review*, 96(2):270–274.
- Kerr, W. R. and Mandorff, M. (2015). Social Networks, Ethnicity, and Entrepreneurship. Technical report, National Bureau of Economic Research.
- Li, Y., Liu, H., Tang, Q., Lu, D., and Xiao, N. (2014). Spatial-Temporal Patterns of China’s Interprovincial Migration, 1985-2010. *Journal of Geographical Sciences*, 24(5):907–923.
- Long, W., Appleton, S., and Song, L. (2017). The Impact of Job Contact Networks on Wages of Rural-Urban Migrants in China: A Switching Regression Approach. *Journal of Chinese Economic and Business Studies*, 15(1):81–101.
- Loury, L. D. (2006). Some Contacts Are More Equal than Others: Informal Networks, Job tenure, and Wages. *Journal of Labor Economics*, 24(2):299–318.
- McKenzie, D. and Rapoport, H. (2007). Network Effects and the Dynamics of Migration and Inequality: Theory and Evidence from Mexico. *Journal of Development Economics*, 84(1):1–24.
- Mincer, J. (1958). Investment in Human Capital and Personal Income Distribution. *Journal of Political Economy*, 66(4):281–302.

- Munshi, K. (2003). Networks in the Modern Economy: Mexican Migrants in the US Labor Market. *The Quarterly Journal of Economics*, 118(2):549–599.
- Patel, K. and Vella, F. (2013). Immigrant Networks and Their Implications for Occupational Choice and Wages. *Review of Economics and Statistics*, 95(4):1249–1277.
- Ye, J., Bo, S., Cong, L., and Zhou, Y. (2012). The Levels of Social Network and Wage of Rural-Urban Migrants: From a Perspective of Identity Model. *Economic Review*, 4.
- Yogo, U. T. (2011). Social Network and Wage: Evidence from Cameroon. *Labour*, 25(4):528–543.
- Zhang, J. and Zhao, Z. (2015). Social-Family Network and Self-Employment: Evidence from Temporary Rural-Urban Migrants in China. *IZA Journal of Labor & Development*, 4(1):4.
- Zhang, X. and Li, G. (2003). Does Guanxi Matter to Nonfarm Employment? *Journal of Comparative Economics*, 31(2):315–331.
- Zhang, Y., Mouhoud, E., and Fan, Y. (2012). Heterogeneous Social Network and Rural Migrant Worker’s Wage: Evidence from China. *South China Journal of Economics*, 2.
- Zhao, Y. (2003). The Role of Migrant Networks in Labor Migration: The Case of China. *Contemporary Economic Policy*, 21(4):500–511.