The Size of Individual Social Networks and Rural-Urban Migrants' Wages—Findings from a Survey of Migrants in China's 6 Provinces and 12 Cities

Renkun Yang, Ruimin Wang, Ran Tao*

Based on a survey of migrants in 12 cities across four major urbanizing areas in China, this paper empirically studies the impact of the size of individual social networks on the migrants' wages. After controlling for potential endogeneity using an instrumental variable approach, our empirical results from 2SLS estimation provides no evidence for significant average causal effect of network size on wage. A further exploration of quantile regression analysis with endogeneity issue managed by using the control function approach shows that a significant positive network size effect can only be found in the low-income end.

Keywords: rural-urban migrants, China, social networks, wage returns

1. Introduction

Among the literature of labor economics, roles of social capital and social networks in the labor market have grabbed considerable interest. One prominent example of how social networks play a role is that social networks can significantly facilitate job search. Studies over the recent decades show that throughout the world at least 30%~50% of job seekers got their jobs through the help of relatives or friends (Rees, 1966; Addison and Portugal, 2002; Wahba and Zenou, 2005; Zhang *et al*., 2008; Topa, 2011).

The rural-urban migrants in China is a group in the labor market among which social networks play a particularly important role. The reason is two-folds. First of all, China is traditionally recognized as a "*Guanxi* society" where social connections and personal relationships are in a central position. Though a comprehensive economic reform aiming at establishing a modern market economy is taking place, such a traditional feature is still embedded in the institutional transmission and plays an

* Renkun Yang, Assistant Professor, School of Economics, Jinan University; Ruimin Wang(corresponding author, Email: cauwrm@foxmail.com), Associate Researcher, Institute of Marketing Economy, Development Research Center of the State Council; Ran Tao, Professor, School of Humanities and Social Science, The Chinese University of Hong Kong, Shenzhen. Fund project: the National Natural Science Foundation of China "Research on the Key Factors and Reform of New Urbanization" (71533007) and Major Project of the National Social Science Fund of China "Research on the Institutional Innovation for Farmers Receiving More Land Property Rights" (17ZDA075).

unneglectable role in market transactions[1] (Zhang and Zhao, 2015). On the other hand, compared with natives, the migrant workers as "newcomers" are usually lack of access to job information, and therefore tend to find jobs with the help of relatives, friends and villagers who have been settled in the same city (Borjas, 1992; Munshi, 2003, 2011). These two features of the rural-urban migrants in China justify the importance of social networks for this group. Furthermore, investigating the specific role that social networks play as well as the underlying mechanism is the key for understanding the interaction between the market and the society in the process of economic reform and how the migrant workers make a prominent contribution in the rapid urbanization and economic growth in China.

In this paper, we attempt to identify the effect of individual network size on rural-urban migrants' wages in China. The proxy we use for the size of social networks is "the number of relatives and friends in the city where the migrant works", which limits the social network to the working cities and is more likely to be related to the migrant worker's job seeking and wages. We find that the significant positive correlation between migrant workers' wages and their individual network size is an overestimate due to endogeneity of social networks. Once the endogeneity problem is corrected by the IV approach, there's no evidence supporting a positive mean causal effect of social network size on wage. A further exploration in quantile estimations, on the other hand, shows that although network size has no significant impact at most quantiles above 40%, it does benefit the lower end migrants, which confirms the conjecture that lower-income migrants might rely more on their social networks.

This paper has two major contributions to the literature. Firstly, we introduce the provincial-level geopolitical immigration network, the number of samples from the same province in the destination city, as the instrument for individual social networks to correct potential endogeneity problem. As will be discussed in more details later, many instruments used in the existing literature are not plausible enough in the exclusive condition (probably correlated with unobserved disturbance). Our instrument is not related to the characteristics of unobservable capacity at the individual or family level, and thus corrects the problem of omitting variables. Furthermore, we propose a modified version in order to rule out the potential selection problem at the group level, providing robust evidence that the size has no mean causal effect. Secondly, we investigate the quantile effect of social network size on migrants' wages. To the best of our knowledge there are not many papers on this topic (Grootaert, 1999; Woolcock and Narayan, 2000; Zhou, 2012), and most of the existing ones have not considered the endogeneity problem as we do. Our result based on the control function approach (Lee, 2007) shows that the quantile

[1] Surprisingly, even in typical modern market economy, such as U.S. economy, reliance on friends, acquaintances and other social contacts is preserved rather than perishing in the past few decades, at least in terms of job market (Ioannides and Loury, 2004).

estimations might be misleading without correcting endogeneity: underestimating the coefficients for lower quantiles and overestimating the coefficients for higher quantiles.

The rest of this paper proceeds as follows. Section 2 reviews related literature in both general contexts and in particular China's rural-urban migration. Section 3 introduces the data and reports the descriptive statistics of relevant variables. Section 4 discusses our main empirical strategies. Section 5 reports the empirical results and Section 6 concludes.

2. Literature Review

This paper contributes to a large class of literature in the field of social networks and China's rural-urban migration. Despite the richness of existing literature, there seems not to be a generalized framework, nor a widely accepted agreement on the impact of social networks on labor market outcomes. This is due to three challenges faced by studies in this topic: complexity and conceptual vagueness of the term "social networks" (Portes, 2000; Manski, 2000; Mouw, 2006), richness of underlying mechanisms that's hard to be characterized in rigorous economic models, and difficulty in causal effect identification given the nature of social network variable and limited data resources (Durlauf, 2002).

First of all, the concept of "social networks" is multi-dimensional. According to Nahapiet and Ghoshal (1998), the impact of social networks can be jointly decided by the structural dimension (overall pattern of connections between actors), relational dimension (interaction and intimacy) and cognitive dimension (reciprocal behavior and common values). It is difficult to make a general analysis and a convincing comparison between different empirical results if social network proxies in different dimensions are used (Durlauf and Fafchamps, 2005). In practice, current empirical literature uses different proxies for social network variables and has not reached an agreement on the network effect on wages, nor a general framework to analyze the possible economic mechanism. Some studies find no significant network effects on workers' wages (Mouw, 2003; Liu and Zhang, 2007; Zhang *et al*., 2008; Zhang and Lu, 2009), while some others find a significantly positive effect (Munshi, 2003; Knight and Yueh, 2008; Ye *et al*., 2012; Hensvik and Nordstrom, 2013; Ye and Wu, 2014; Burks *et al*., 2015). Even if we restrict the emphasis on the size of individual social network size, the proxy variables are still different case by case. For example, Munshi (2003) finds that the expansion of the migration network from the same source-community could help Mexican immigrants increase the probability of getting non-agricultural jobs in the United States and also get higher wages. Knight and Yueh (2008) on China rural-urban migrants and Amuedo-Dorantes and Mundra (2007) on Mexican-American immigrants both use "the number of relatives and friends of the family" as a proxy for social networks and find significant impacts on both job seeking and wages. On the other hand, Zhang and Lu (2009) investigate the impact of "the number of relatives and friends to whom an individual has presented gifts", which reflects not only the size of individual or familial social networks but also the willingness to invest in social networks, but find little impact of size on wages.

Secondly, the different dimensions of social networks may have distinct channels to impact outcome variables (Zhang and Li, 2003). In economic literature, most formal theoretical models on the wage effect of social networks focus on three categories: overcoming inefficiency from information asymmetry through job referral, information advantage in job search, and peer effect on productivity. The seminal work by Montgomery (1991, 1992) shows that when heterogeneous worker productivity is private information, the recommenders are willing to act as a screening mechanism for the sake of their own career development, resulting in the employer's willingness to offer more favorable positions to the referred job seekers. Mortensen and Vishwanath (1994) further demonstrate the information effect: even with identical job seekers, those with a higher probability of obtaining offer information through employed contacts tend to earn more due to more favorable wage information.[1] Besides the information effect, Delattre and Sabatier (2007), Ye *et al*. (2012) and Lindquist *et al*. (2015) investigate the productivity effects of peer pressure, showing that high quality co-worker (or acquaintance) social networks can motivate workers to spend more effort and thus obtain higher wages.[2] Despite the richness of recent theoretical research on network effect on wages, most of the existing models focus on referral using or quality of co-worker networks. The impact of network size, on the other hand, is hard to formalize. Wahba and Zenou (2005) develop a model to build a link between the size and the use of social networks. They find that conditional on being employed, the probability of finding a job through social networks increases with the size of the networks, and a survey from Egypt provides empirical evidence supporting their conclusions. In a similar manner, Munshi (2003) uses a statistical model involving referrals (analogous to the Montgomery setting) and confirms that the expansion in network size can improve the labor market outcome of Mexican immigrants in the U.S.

Finally, though some of the earlier studies pay little attention to the endogeneity problem of social networks (Lin, 2001; Mouw, 2003), it has soon become a central

[1]This is because of the classical cutoff rule of the job searching literature: job seekers only accept offers when the offered wage is higher than the reservation. Since the employed contacts are workers who already accepted above-cutoff wages, the wage information from them tends to be skewed to the higher end in distribution. Thus job seekers obtaining offer information from these contacts and finally accepting the offer will in general receive higher pay than others in the sense of first-order stochastic dominance.

[2] As discussed in Lindquist *et al*. (2015), the positive productivity effect of co-workers can also be caused by technology spillovers. They distinguish the peer pressure effect from the spillover effect and provide effective identification for empirical study.

issue in the empirical literature.[1] Instrument variable (IV) approach is the most commonly used identification strategy with three types of instruments: natural conditions (Munshi, 2003; Zhang and Zhao, 2015), family political background (Zhang and Lu, 2009) and the social and cultural background (Narayan and Pritchett, 1999; Ye and Zhou, 2010; Ye and Wu, 2014). Munshi (2003) uses the lagged rainfall of the migrants' origin as an instrument for the Mexican-American community migration networks. Such exogenous shocks effectively satisfy the exclusiveness and correlation conditions while imposing strong data and background limitations on external applications. Zhang and Zhao (2015) use the distance between the capital city of the home province and the working city as the instrument to study the impact of the size of relatives and friends on the migrants' choices of self-employment, but the instrument is difficult to apply to the study on wages because of the migrating self-selection bias brought by omitting variables.[2] In terms of the families' political, social and cultural backgrounds, Zhang and Lu (2009) use the political classes of the spouse family during the Cultural Revolution as an instrument for the number of relatives and friends. But family political classes are often related with economic background, genetic capacity, family education and other unobservable variables. Even the spouse's family background is also related if we take the assortative mating into account. Ye and Wu (2014) use the Spring Festival greetings as an instrument by arguing that this behavior is more influenced by customs and could been seen as exogenous. However, it is difficult to rule out the possibility of inverse causality: migrants with high income are more likely to give a New Year's call than those with low income due to vanity concern.

3. Data and Descriptive Statistics

In 2009, one of the authors conducted a large-sample migrants survey in 12 cities across four major urbanized regions: the Yangtze River Delta (Jiangsu and Zhejiang Province), the Pearl River Delta (Guangdong province), the Chengdu-Chongqing region (Sichuan Province and Chongqing Municipality), and the Bohai Bay Area (Hebei and Shandong Province). We randomly select one megalopolis (urban population larger than 2 million), one large city (500000~2 million), and one small- and medium-sized city (<500000) in each of the four urbanized regions. Due to the huge number of migrants in the megalopolis, we only sample in one urban district randomly selected in the megalopolis, while in the large and small- and medium-sized cities we target all urban districts in the city as our sampling frame. Then we randomly select 200 migrants in each city (2400 migrants in total) from the migrant registration

[1] See Chen and Fan (2011) for a comprehensive survey on causal identification in social network literature.

[2] Zhang and Zhao (2015) pointed out that this problem may also exist in the studies of self-employment selection.

list provided either by the local Public Security Bureau or by local government migrant administrative agency.[1] Migrants are defined as people whose Hukou is not registered in the city in which they lived at the time of survey, and they have left their Hukou registration places for more than three days.[2] Finally, 2299 valid samples were obtained. The survey collected information about migrants' various aspects of livelihood, including demography, employment, income, social network, and so on.

The main variables in this study include wage, social network, human capital and other personal characteristics. Table 1 reports the descriptive statistics of the main variables. The dependent variable is the logarithmic hourly wage of the migrant worker in 2008, and the average hourly wage is 7.39 yuan (RMB). 56.6% of the samples are male, the average age is 32 years-old, and 62% of the samples are married. In terms of social networks, including the number of friends/relatives in the same city and the total number, a migrant worker averagely has 8 friends and 4 relatives in the city where they work. In terms of human capital, the migrants' average years of schooling is 9 years (roughly equivalent to junior high school) and the working experience in the current city is 7 years on average. In addition, the proportion of CPC party members is only 6%, which is consistent with our experience and findings in the literature.

Table 2 reports the statistics of wages and the size of social networks of migrant workers in the four major urbanized regions. It can be seen that the average wage of the migrant workers in the Bohai Bay Area are relatively low and the size of social networks is also smaller, while the migrant workers of the Pearl River Delta have the highest average wage and the largest network of friends and relatives.

Table 1. Descriptive Statistics of the Main Variables

Variables		Obs	Mean	Std	Min	Max
Dependent variable	Hourly wage	2295	7.39	8.70	0.324	167
Social networks In the same city	Number of relatives and friends	2093	12.73	20.29	0	280
	Number of friends	2194	8.16	17.18	0	210
	Number of relatives	2127	4.30	7.05	0	120
Human capital	Education	2299	8.96	3.48	0	18
	Experience	2299	7.07	6.29	0	40
	Square of experience	2299	89.5	144	0	1600

[1] Not all migrants are registered. The registration rate ranges from about 70% of migrants in Guangdong province to 90% in Jiangsu and Zhejiang province. If the sampled migrant has already moved away, we continue the systematic random sampling until we reach the desired sample size. The percentage of replacement due to migrant unavailability in sampling ranges between 15% to 30% in different cities, usually higher in larger cities.

[2] The current policy is that migrants who have left their hometowns for more than 3 days are required to apply for a temporary residence permit.

Variables		Obs	Mean	Std	Min	Max
Demographics	Gender: male=1	2299	0.565	0.496	0	1
	Age	2299	31.9	10.1	16	81
	Marital status: married=1	2297	0.620	0.486	0	1
	Party: CPC member=1	2298	0.0596	0.237	0	1
	Minority: Han =1	2298	0.949	0.220	0	1

Table 2. Wages and Social Networks in Four Major Urbanized Regions

	Yangtze River Delta			Pearl River Delta			Bohai Bay Area			Chengdu-Chongqing region		
Varible	Obs	Mean	Std.	Obs	Mean	Std.	Obs	Mean	Std.	Obs	Mean	Std.
Hourly wage	569	7.42	6.78	568	8.61	10.51	587	6.63	6.27	571	6.94	10.30
Number of relatives and friends	460	10.4	14.4	554	16.84	26.15	517	9.40	19.55	562	13.61	17.55
Number of friends	509	5.55	10.5	560	11.21	21.90	558	1.11	1.22	567	8.99	15.03
Number of relatives	475	4.47	7.00	563	5.52	9.13	524	2.56	4.49	565	4.55	6.32

4. Empirical Strategy

4.1. Model Specification

In this paper, we use the classic Mincer wage equation as the basic model:

$$\ln hourwage = X\alpha + N\beta + \varepsilon \tag{1}$$

The dependent variable is the logarithm of migrant worker's hourly wage. The explanatory variable X is a vector of control variables including gender, marital status, education, city work experience, square of work experience, party, minority and constant term, among which gender (male=1), marital status (married=1), party (party member=1) and minority (minority=1) are dummy variables. Home province and destination city fixed effects will also be included in further model specifications. N is the proxy variable for the size of the individual social networks, including the logarithm of total number of friends and relatives. ε is the error term.

In this paper, we first run an OLS regression to estimate the impact of the size of social network on wages. But the consistency and unbiasedness of the estimation depend on the exogeneity of the explanatory variable, i.e., $E(N\varepsilon)=0$. However, it is probable that this condition is violated. In this case, it is necessary to find an effective identification strategy to make causal inference. Here we take the standard IV approach and introduce an instrument which affects the wage only through individual social networks. Following is the first-stage regression equation for social network determination:

$$N = X\gamma_1 + IV\gamma_2 + U \tag{2}$$

Here X is the vector of control variables, U is the error term and IV (the aggregate level geopolitical immigration network) is our instrument. Given that assumption $E(IV\varepsilon)=0$ holds which we'll discuss in more details in the next subsection, we can perform a two-stage least square regression to get a consistent estimation for the network effect.

The third set of specification is the quantile regressions. Since the network size variable might incur the endogeneity problem, the standard quantile approach might lead to the same problem with OLS. Thus we follow the Control Function Approach (CFA) by Lee (2007) to perform the quantile regressions while dealing with the endogeneity problem. The model setting is the same as above, while the estimation follows a two step control function procedure. The first step is carried out by a linear mean regression (OLS) of individual network size on the instrument and controls.[1] The residual for the individual network equation is obtained by:

$$\hat{U} = N - X\hat{\gamma}_1 - IV\hat{\gamma}_2 \tag{3}$$

Then the second step is the quantile regression of Ln*hourwage* on N, X and $\hat{U}$. Lee (2007) has shown that the estimator obtained from this two-step procedure is $n^{-1/2}$-consistent and asymptotically normal under a minimal set of regulatory conditions.

4.2. Instrument Choice

In this paper, we use the size of migrant networks at the provincial level as an instrument for the size of the social networks at the individual level. The proxy can be described as "number of workers from the same province working in the same city" in our data set. It's well known that a desirable instrument needs to satisfy two conditions: correlation and exclusion. In terms of correlation, this instrument represents the number of potential fellows in the city, and the correlation with the size of the individual social networks is intuitive: when there are more migrant workers from the same home province there will not only be more friends and relatives in the same destination city for an individual, but also more opportunities for him/her to expand the social networks based on the same geographical and cultural background. The empirical

[1] Lee (2007) analyzes a general triangular equations model where the first step could be a linear quantile or mean estimation for the construction of residuals of the reduced-form equations for the endogenous explanatory variable. See Lee (2007) for examples of both linear quantiles first step (demand for fish) and linear mean first step (returns to schooling).

result from the first stage regression further confirms the correlation restriction.

More importantly, we argue that the migration network, which represents the size of potential "fellows", has no direct impact on the individual wage, and is hard to be directly affected by the individual wage either. Thus it should satisfy the exclusion restriction of the instrument. It is worth noting that this instrument is very similar to the variable of "community migration network" which is commonly used in the literature of labor migration. However, there are some important distinctions. Social migrant network variables are often described in the literature as "the number or proportion of people migrating outward",[1] constructed on the basis of origin village or even smaller unit of community (Massey *et al*., 1994; Zhao, 2003; Mckenzie and Rapoport, 2010; Munshi, 2003, 2011). These variables usually do not only affect individual wages through individual social networks, but are also potentially associated with the unobserved heterogeneity of small groups such as clans, communities, and villages, and are therefore not suitable as instruments for the size of individual networks. Our instrument of migration network, on the other hand, represents the "density of same-origin-province migrants in the destination" rather than the traditional "density of outwards migrants in the origin", where the province-level based variable is much less overlapped with individual social networks and have no correlation with the omitting variables at the level of clans, communities or villages.

Another potential problem is that the instrument, while not directly related to the individual omitting variables, may be indirectly related to the wages through the self-selection process of migration decision.[2] In more details, individuals with lower level of unobserved ability usually migrate to the cities close to the origin, and thus gather in neighboring cities to form a large network. On the contrary, migrants with higher skills are capable of migrating to distant cities, so their destinations are relatively scattered and the home-province based networks are smaller. This idea coincides with Munshi's (2011) discussion of the bias estimates for social network variables: when a positive selection bias of abilities exists, that is, those with higher ability tend to migrate, the expansion of the social networks will lower the marginal abilities, and therefore social networks and the residual term will be negative related. To the opposite, when the selection bias is negative, the explanatory variables and residuals are positively related.

In order to minimize such estimation bias, we first add the "migration distance" as a control variable in the two-stage least squares model to check whether the network effect estimation will be affected. This is due to the concern that migrant workers with low ability are clustered to form a large migration network because they are incapable of covering the high migration costs associated with distance, and thus the potential selection problem

[1] In more details, in most studies of Mexican-American migration, the measurement of the migration network is "the proportion of people who have moved out of the residents aged 15 years and over in a given community".

[2] We are thankful to professor Christopher Taber's for pointing this out.

of the instrumental variable becomes weaker after controlling distance. Here we use two proxies for distance. One is “working in the hometown/ the county outside the hometown/ the province outside the home county or other provinces”. This distance variable actually represents the administrative and cultural distance, and pays more attention to the short-range migration. The second variable is “the distance between the current city and the capital city of the home province”[1] which represents the natural geographical distance. It is more accurate to describe the long-distance migration, but to the migrant workers in the home or neighbor provinces, the measurement error may be larger.

Another improvement is to modify the original instrument, that is, to “peel” the potential selection bias caused by the migration cost away from the instrument. Based on this idea, we first regress the original instrument on the distance variables, and then get the predicted residual as the new instrument. This instrument will outperform the original one if the selection issue exists. The only concern is that the modification might lead to weak IV problem. As is shown in Section 5, this concern is not a problem since the modified instrument is not much weaker than the original one, and the estimation result is supported by Anderson-Rubin test which is robust to potential weak IV risk.

5. Empirical Results

5.1. OLS and IV 2SLS Results

First we perform the OLS estimation. The dependent variable used in all the five specifications is the logarithm of the hourly wage, and the core explanatory variable is the size of individual social networks. Table 3 reports the OLS results, where Columns (1) to (3) examine the effect of total network size. Specification 1 in the first Column investigates the effect of total number of friends and relatives (logarithm) in the same city. Based on this specification, Column (2) controls the fixed effects of destination city and Column (3) further controls the fixed effects of the home provinces. Since both the dependent variables of the model and the social networks variables are logarithmic, the coefficients of the social network variables actually represent the hourly wage elasticity of the networks. From the regression results in the first column,we can see that the hourly wage elasticity of the total social network size is 0.0244 and is statistically significant at 5% level. After controlling the destination city fixed effects, the coefficient of network size decreases to 0.0194, while remaining significant at 10% level. Other coefficients have no systematic change.We further control the home-provincial fixed effects in Column (3). The hourly wage elasticity of the networks is still significant, and the coefficients of the regression equation again do not change systematically.

[1] This variable coincides with the instrument of Zhang and Zhao (2015). However, due to the selection bias argument here, we don’t treat distance itself as a plausible instrument.

We examine the effects of dummies indicating the destination cities and home provinces.[1] We find that destination cities have a significant impact on wages but the origin provinces make little difference, which is consistent with the economic intuition and the existing literature (Liu and Lin, 2007).[2] Therefore, we will only consider the fixed effects of the destination cities in further analysis.

Table 3. Wage Returns of Network Size: OLS Regression (Dependent Variable: ln (Hourly Wage))

Variable	Model 1	Model 2	Model 3
ln (number of relatives and friends in the same city)	0.0244**	0.0194*	0.0224**
	(0.0101)	(0.0100)	(0.0100)
Male=1	0.127***	0.135***	0.133***
	(0.0226)	(0.0222)	(0.0223)
Education	0.0522***	0.0531***	0.0516***
	(0.00378)	(0.00379)	(0.00387)
Experience	0.0324***	0.0285***	0.0291***
	(0.00530)	(0.00502)	(0.00507)
Square of experience	−0.000840***	−0.000599***	−0.000613***
	(0.000219)	(0.000201)	(0.000206)
Married=1	0.0221	0.0143	0.00320
	(0.0287)	(0.0289)	(0.0292)
CPC member=1	0.0956*	0.0969*	0.0981*
	(0.0555)	(0.0554)	(0.0554)
Minority: Han=1	−0.107**	−0.0586	−0.130**
	(0.0493)	(0.0491)	(0.0541)
Destination city fixed effect	No	Yes	Yes
Home province fixed effect	No	No	Yes
Constant	1.272***	1.265***	1.414***
	(0.0678)	(0.0758)	(0.106)
Obs	2,088	2,088	2,088
R^2	0.144	0.185	0.206

Note: The robust standard error for the estimated coefficient is in the parentheses, *** p <0.01, ** p <0.05, * p <0.1. All the models in Table 3 are OLS regressions with double logarithm setting. Models 1 and 2 do not control the area fixed effect, while models 3 and 4 control the fixed effect of the city, and models 5 and 6 further control the fixed effect of the home province.

[1] Due to limitation of pages, estimated coefficients are not presented in Table 3.

[2] Taking specification 3 as an example, the coefficients of seven variables in 11 urban dummy variables are significant, with Ningbo city as the control. The wages of migrant workers in Wenzhou and Dongguan are significantly higher than wages in Ningbo, and the wages in Jinan, Langfang, Weifang, Chongqing, Nanchong are relatively lower. In contrast, only 4 coefficients out of the 28 provinces are significant, and the number of migrant workers from these four provinces is also less than 5% of the total sample. It can be concluded that the impact of the provincial dummies is very small.

5.2. IV 2SLS Results

Panel A in Table 4 reports the results of the two-stage least squares estimation based on the instrument choices, where Columns (1)~(3) are results using the original provincial migration network as the instrument and Columns (4) and (5) use the modified instrument that "peels" out potential selection. Column (1) shows the estimation results with the same set of controls as the OLS regression. We find that the coefficient of social networks becomes negative but statistically insignificant. Meanwhile wage gap in gender and return to work experience become larger, and the coefficients of minority and CPC membership also change. This result indicates that the size of social networks has no significant effect on the wage of migrant workers after correcting endogeneity problem. As discussed in Section 4, there might be potential selection problem with the original instrument. To roughly test and control for this effect, we add "the railway distance" and "administrative distance" as control variables in Columns (2) and (3). The results suggest that the absolute value of coefficient of social network decreases and remains insignificant. This confirms the plausibility of the modified instrument over the original one. Columns (4) and (5) reports the results of the 2SLS estimator based on the modified instrument. It can be seen that the absolute value of the coefficient of the number of relatives and friends is also reduced and still not significant. All these results suggest that the positive impact of the size of social networks on wages in OLS disappears after taking endogeneity into consideration, and it is more likely a correlation relationship resulting from endogeneity rather than causality.

Panel B in Table 4 reports the first stage regression results. From the first stage coefficient it can be seen that both instruments are highly correlated with the individual network size (significant at 1% level in all five models) which confirms the correlation condition for a plausible instrument. We further examine weak IV test. It's well known that the unbiasedness and consistency of the coefficient estimates might be violated if the correlation between the instruments and the key explanatory variables is very weak (Stock *et al*., 2012). According to the rule of thumb proposed by Staiger and Stock (1997), a frequently used rule of thumb that indicates a potential weak IV problem is an F-statistics smaller than 10. In panel B, we can find that all the F-statistics from the 5 settings are slightly below 10, which means that our results might suffer from weak IV problem. Fortunately, we can confirm the conclusion here using the robust weak IV test approach proposed by Stock and Wright (2000). This seminal research on weak IV indicates that we can rely on the Anderson-Rubin (AR) test to test the coefficient estimates as well as get robust interval prediction, even if the point estimation from 2SLS regression is under the risk of weak IV. Panel C in Table 4 shows the AR test result with the null hypothesis "network size coefficient = 0". P-value from all five specifications are sufficiently large such that we cannot reject the hypothesis. This confirms the conclusion from the 2SLS estimation that the network size has no significant impact on migrants' wages, once endogeneity problem is corrected. The estimated 95% confidence intervals also suggest that the true coefficient should lie somewhere around 0.

Table 4. Wage Returns of Network Size: IV 2SLS Regression (Dependent Variable: ln (Hourly Wage))

Panel A					
	Second stage				
Variable	Original IV:			Modified IV:	
	Model (1)	Model (2)	Model (3)	Model (4)	Model (5)
ln (number of relatives and friends)	−0.227	−0.0113	−0.0104	0.0343	0.0366
	(0.187)	(0.144)	(0.144)	(0.150)	(0.150)
Male=1	0.143***	0.133***	0.133***	0.135***	0.133***
	(0.0271)	(0.0232)	(0.0232)	(0.0230)	(0.0232)
Married=1	−0.000808	0.00622	0.00720	0.0116	0.0103
	(0.0342)	(0.0292)	(0.0292)	(0.0293)	(0.0293)
Education	0.0516***	0.0530***	0.0531***	0.0536***	0.0530***
	(0.00423)	(0.00363)	(0.00363)	(0.00366)	(0.00362)
Experience	0.0387***	0.0287***	0.0286***	0.0262***	0.0265***
	(0.0105)	(0.00841)	(0.00842)	(0.00878)	(0.00864)
Square of experience	−0.000905***	−0.000627**	−0.000625**	−0.000567**	−0.000577**
	(0.000339)	(0.000278)	(0.000278)	(0.000286)	(0.000282)
CPC member=1	0.0309	0.0688	0.0685		0.0785
	(0.0661)	(0.0553)	(0.0553)		(0.0558)
Han=1	−0.0508	−0.0691	−0.0692		−0.0788
	(0.0622)	(0.0529)	(0.0529)		(0.0531)
ln (geographical distance)[a]		0.00376**	0.00378**		
(Railway)		(0.00189)	(0.00189)		
Geographical distance [b]			−0.000597		
(administrative)			(0.000555)		
City fixed effect	Yes	Yes	Yes	Yes	Yes
Constant	1.659***	1.299***	1.300***	1.197***	1.273***
	(0.286)	(0.228)	(0.228)	(0.243)	(0.231)
Obs	2,042	2,042	2,042	2,042	2,042
R^2		0.183	0.184	0.184	0.186
Panel B					
	First stage				
Original instrument	0.0638**	0.0950***	0.0950***		
	(0.0254)	(0.0319)	(0.0319)		
Modified instrument				0.0897***	0.0899***
				(0.0312)	(0.0312)
F-statistics	8.39	8.16	9.24	9.31	8.54

Variable	Second stage				
	Original IV:			Modified IV:	
	Model (1)	Model (2)	Model (3)	Model (4)	Model (5)
			Panel C		
	Anderson−Rubin test (H0: β=0)				
p-value	0.20	0.96	0.95	0.71	0.70
95% confidence set	[−1.02, 0.12]	[−0.33, 0.39]	[−0.32, 0.37]	[−0.28, 0.51]	[−0.28, 0.52]

Note: The robust standard error for the estimated coefficient is in the parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.
a: The variable is “ln (distance between the current city and the capital city of the home province)”, it represents the natural geographical distance.
b: This variable is based on the question of the questionnaire. “Your present employment location is: 1 village, 2 the town outside your home village, 3 the county outside your home town, 4 the province outside your home county, 5 other provinces,6 foreign country” and can be regarded as the distance variable in the term of administration. The samples in this paper are mainly concentrated in the province outside the home county or other provinces.

5.3. Quantile Regression Results

At last we perform the quantile regression results. Figure 1 shows the coefficient estimates of total network size as well as the confidence intervals at 90 quantiles from 0.05 to 0.95. The upper figure is the result from quantile regressions without considering endogeneity problem, while the lower figure takes network size to be endogenous and uses control function approach proposed by Lee (2007). In this section we use the modified instrument to perform the first stage regression, which is no different from the first stage in the last section. Then the predicted residual is used as one of the control variables to perform quantile regressions. The first observation is that at high wage quantiles (above 0.8) estimations in both specifications are highly chaotic. This means that for the most successful rural-urban migrants, individual network size has no explanatory power. The second observation is that the coefficients in the first specification are steady around 0.025, which is quite similar to the OLS estimation. Meanwhile coefficients in the CFA specification are in general decreasing along quantiles from positive values to negative. To illustrate the significance level of the quantile regression, we present coefficients for below-mean quantiles in Table 5. It’s clear that at the lower end (below 25%), social network size plays a significantly important role to help migrant workers get higher wage. In contrast, for above-mean quantiles, we cannot obtain coefficients that are significantly different from 0.

The results show that the wage return to network size would be underestimated for lower quantile migrants and overestimated for higher quantile migrants, if we don’t consider endogeneity. This means that lower income migrants might rely more on their social networks than higher income migrants. One possible interpretation is that social networks might be substitutes to human capital and unobservable working abilities. Lower income migrant workers are usually associated with lower unobserved ability and have less capability of collecting job information. It's more difficult to obtain a higher wage job position by their own means, thus they rely more on the social networks they have. While the case is much different for higher ability migrant workers, even though they are more likely to obtain more and better social capital. Some of them might rely heavily on a larger number of friends or relatives to get more income, especially the successful self-employed migrants, while some others get the high-pay job through their high productivity in a remote city from their hometown where they have few friends or relatives to ask for help.

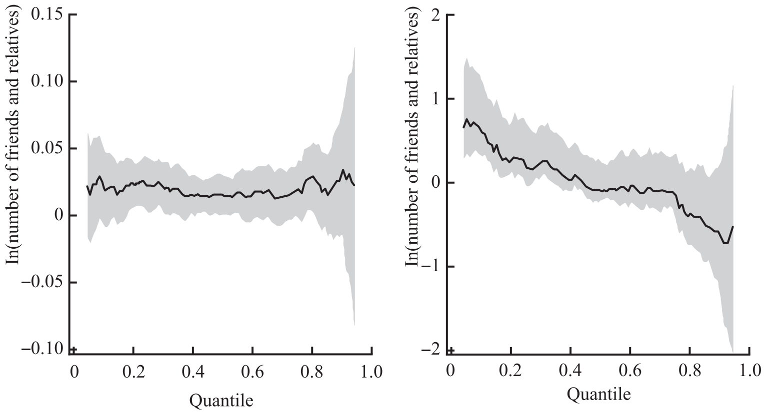

Figure 1. Coefficients of Network Size at 0.05~0.95 Quantiles

Table 5. Coefficients of Network Size at Below-Median Quantiles

Quantiles	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
ln (total number)	0.701*	0.684***	0.397*	0.278	0.306*	0.242	0.186	0.0526	0.00859	−0.0741
	(0.359)	(0.233)	(0.203)	(0.184)	(0.167)	(0.159)	(0.142)	(0.134)	(0.130)	(0.132)

6. Conclusions

Based on a survey of migrants in China's 6 provinces and 12 cities, this paper empirically studies the impact of the size of individual social networks on the migrant workers' wages. Our identification strategy is to use the aggregate migration network as the instrument for individual social networks. Based on the empirical results from IV 2SLS estimations and quantile estimations with endogenous explanatory variable, this paper draws the following conclusions. Firstly, the size of individual social networks and wages are significantly correlated. Secondly, after dealing with the endogeneity problem by IV approach, we cannot find evidence for significant wage effect of network size. Anderson-Rubin robust test further confirms this conclusion, no matter the threat of weak IV exists or not. Our result in general agrees with the conclusion of Zhang and Lu (2009), but differs from most other existing studies that find positive network size effect on wages (Munshi, 2003; Knight and Yueh, 2008). Thirdly, a further exploration in quantile estimations based on control function approach shows that at lower wage quantiles, migrants can benefit from larger individual networks, while this doesn't hold for migrants at higher wage quantiles.

It's worth noting that the insignificant results in this paper do not imply that social networks have little effect on the performance of migrants in the urban labor market. Firstly, the key network variables used in this study do not include the migrants' contacts among urban natives, nor the "neighborhood" and "community" that are valued in the literature of social interaction and peer effects. Secondly, the size is only one dimension of social networks. It's probable that the strength and quality of social networks are playing an important role in the labor market performance, and whether to use social networks to find a job may also matter. Thirdly, wage is only one labor market outcome. Effects of social networks on other variables such as referral using, turnover rates, social integration and length of unemployment are worthy of further investigation.

So, we believe that further studies (especially in China's labor market) could focus on the following directions. Firstly, it's valuable to investigate the impact of social networks on the selection patterns of unobservable ability and observable education in the migration (Mckenzie and Rappoport, 2007). The efforts on this issue will help us better understand the impact of social networks on individuals as well as the income inequality in both rural and urban areas caused by the migration. Secondly, it's also promising to study the investment choice and accumulation process of social capital as well as their interaction with individual income and economic status. The existing studies, especially in sociology, pay a lot of attention on acts such as gift giving or Spring Festival greetings as ways of social capital investment. Further effort should be integrating these network investment behaviors into formal economic models based on the fruitful area of social interaction literature. Thirdly, it might be investigating mechanism of the network effect in depth making use of solid theory foundation and structural estimation. Theoretical literature such as Montgomery (1991) have developed insightful models in early years, while the empirical breakthrough has not appeared until the past decade (Cingano and Rosolia, 2012; Kramarz and Skans, 2014; Schmutte, 2015; Brown *et al*., 2016; Hensvik and Nordstrom, 2016). Most of this literature is based on company-employee matching data, and can be helpful in understanding the role of job referrals on the probabilities of job acceptance, wages, turnover rates, productivity, and welfare implications. More works on this type of research with corresponding data-set in China are in need to more precisely and rigorously explain the socio-economic effect of social networks in both urban labor market and rural-urban groups.

References

Amuedo-Dorantes, C., & Mundra, K. (2007). Social Networks and Their Impact on the Earnings of Mexican migrants. *Demography*, 44(4), 849–863.

Addison, J. T, & Portugal, P.(2002). Job Search Methods and Outcomes. *Oxford Economic Papers,* 54(3), 505–533.

Borjas, G. J., Bronars, S. G., & Trejo, S. J. (1992). Self-Selection and Internal Migration in the United States. *Journal of Urban Economics*, 32(2), 159–185.

Brown, M., Setren, E., & Topa, G. (2016). Do Informal Referrals Lead to Better Matches? Evidence from a Firm's Employee Referral System. *Journal of Labor Economics*, 34(1), 161–209.

Burks, S. V., Cowgill, B., Hoffman, M., & Housman, M. (2015). The Value of Hiring through Employee Referrals. *Quarterly Journal of Economics*, 130(2), 805–839.

Chen, Y., & Fan, X. (2011). Measuring the Labor Market Effects of Social Capital: A Literature Review and Research Strategy of Dealing with the Endogeneity Problem. *Sociological Studies (Shehuixue Yanjiu)*, 1,167–195.

Cingano, F., & Rosolia, A. (2012). People I know: Job Search and Social Networks. *Journal of Labor Economics*, 30(2), 291–332.

Delattre, E., & Sabatier, M. (2007). Social Capital and Wages: An Econometric Evaluation of Social Networking's Effects. *Labour*, 21(2), 209–236.

Durlauf, S. N. (2002). On The Empirics Of Social Capital. *The Economic Journal*, 112(483), 459–479.

Durlauf, S. N, Fafchamps, M. (2005). Chapter 26 — Social Capital, in *Handbook of Economic Growth*. Elsevier B.V. 1639–1699.

Grootaert, C., Oh, G., Swamy, A. (1999). Social Capital, Household Welfare and Poverty in Burkina Faso. Policy Research Working Paper, 11(1), 4–38.

Hensvik, L, & Skans, O. N. (2016). Social Networks, Employee Selection and Labor Market Outcomes. *Journal of Labor Economics*, 34(4), 825–867.

Knight, J., & Yueh, L. (2008). The Role of Social Capital in the Labour Market in China. *Economics of transition*, 16(3), 389–414.

Kramarz, F., & Skans, O. N. (2014). When Strong Ties Are Strong: Networks and Youth Labour Market Entry. *Review of Economic Studies*, 81(3), 1164–1200.

Lee, S. (2007). Endogeneity in Quantile Regression Models: A Control Function Approach. *Journal of Econometrics*, 141(2),1131–1158.

Lin, N., Cook, K. S., & Burt, R. S. (2001). *Social Capital: Theory and Research*. Transaction Publishers.

Lindquist, M. J., Sauermann, J., & Zenou, Y. (2015). Network Effects on Worker

Productivity. CEPR Discussion Paper Series.
Liu, L., & Zhang, C. (2007). Human Capital, Social Capital, Enterprises Institution or Social Environment: Wage Determination Model of Migrant Workers in Peal River Delta. *Sociological Studies (Shehuixue Yanjiu)*, 6,114–137.
Manski, C. F. (2000). Economic Analysis of Social Interactions. NBER Working Paper, No. w7580.
Massey, D. S., Goldring, L., & Durand, J. (1994). Continuities in Transnational Migration: An Analysis of Nineteen Mexican Communities. *American Journal of Sociology*, 99(6), 1492–1533.
McKenzie, D., & 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.
Montgomery, J. D. (1991). Social Networks and Labor-Market Outcomes: Toward an Economic Analysis. *The American Economic Review*, 81(5), 1408–1418.
Montgomery, J. D. (1992). Job Search and Network Composition: Implications of the Strength-of-Weak-Ties Hypothesis. *American Sociological Review*, 57(5), 586–596.
Mortensen, D. T., & Vishwanath, T. (1994). Personal Contacts and Earnings: It Is Who You Know! *Labour Economics*, 1(2), 187–201.
Mouw, T. (2006). Estimating the Causal Effect of Social Capital: A Review of Recent Research. *Annual Review of Sociology*, 32, 79–102.
Mouw, T. (2003). Social Capital and Finding a Job: Do Contacts Matter? *American Sociological Review*, 68(6), 868–898.
Munshi, K. (2003). Networks in the Modern Economy: Mexican Migrants in the US Labor Market. *The Quarterly Journal of Economics*, 118(2), 549–599.
Munshi, K. (2011). Strength in Numbers: Networks as a Solution to Occupational Traps. *The Review of Economic Studies*, 78(3), 1069–1101.
Nahapiet, J., & Ghoshal, S. (1998). Social Capital, Intellectual Capital, and the Organizational Advantage. *Academy of Management Review*, 23(2), 242–266.
Narayan, D., & Pritchett, L.(1999). Cents and Sociability: Household Income and Social Capital in Rural Tanzania. *Economic Development and Cultural Change*, 47(4), 871–897.
Portes, A. (2000). Social Capital: Its Origins and Applications in Modern Sociology. *LESSER, Eric L. Knowledge and Social Capital. Boston: Butterworth-Heinemann*, 43–67.
Rees, A. (1966). Information Networks in Labor Markets. *American Economic Review*, 56(1/2), 559–566.
Schmutte, I. M. (2015). Job Referral Networks and the Determination of Earnings in Local Labor Markets. *Journal of Labor Economics*, 33(1), 1–32.
Staiger, D. O., & Stock, J. H. (1997). Instrumental Variables Regression with Weak Instruments. *Econometrica*, 65(3), 557-586.
Stock, J. H., & Wright, J. H. (2000). GMM with Weak Identification. *Econometrica*,

68(5), 1055–1096.

Stock, J. H., Wright, J. H., & Yogo, M. (2012). A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments. *Journal of Business & Economic Statistics*, 20(4), 518–529.

Topa, G. (2011). Labor Markets and Referrals. *Handbook of Social Economics* (I), 1193–1221.

Wahba, J., & Zenou, Y. (2005). Density, Social Networks and Job Search Methods: Theory and Application to Egypt. *Journal of Development Economics*, 78(2), 443–473.

Woolcock, M., & Narayan, D. (2000). Social Capital: Implications for Development Theory, Research, and Policy. World Bank Research Observer, 15(2), 225–249.

Ye, J., & Wu, L. (2014). The Effect of Social Capital on Migrant Workers Wage Level: Resource Measurement and Identification Strategy. *China Economic Quarterly (Jingjixue Jikan)*, 4, 1303–1322.

Ye, J., Bo, S., Liu, C. *et al*., (2014). The Levels of Social Network and Wage of Rural-Urban Migrants: From a Perspective of Identity Model. *Economic Review (Jingji Pinglun)*, 4, 31–42.

Ye, J., & Zhou, Y. (2010). Social Capital Transition and Income of Rural-Urban Migrants: Evidence from a Survey in Beijing. *Management World (Guanli Shijie)*, 10, 34–46.

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), 1–21.

Zhang, X., & Li, G. (2003). Does Guanxi Matter to Nonfarm Employment. *Journal of Comparative Economics*, 31(2), 315–331.

Zhang, Y., Li, R.,Wang, H. *et al*. (2008). Social Networks and Wage: Empirical Analysis Based on Migrant Worker. *World Economic Papers (Shijie Jingji Wenhui)*, 6, 73–84.

Zhang, Y., & Lu, M. (2009). Does Social Network Help in Improving Wage of Migrant Workers? *Management World(Guanli Shijie)*, 3, 45–54.

Zhao, Y. (2003). The Role of Migrant Networks in Labor Migration: The Case of China. *Contemporary Economic Policy*, 21(4), 500–511.