Network Effects on Labor Contracts of Internal Migrants in China - A Spatial Autoregressive Model

Badi H. Baltagi, Ying Deng and Xiangjun Ma*

February 15th, 2016

Abstract

This paper studies the fact that 40 percent of internal migrants in China do not sign a labor contract with their employers, as revealed in a nationwide survey. These contract-free jobs pay lower hourly wages, require longer weekly work hours, and provide less insurance or on-the-job training than regular jobs with contracts. This paper finds that the co-villager networks play an important role in a migrant's decision on whether accepting such insecure and irregular jobs. In fact, other migrants' propensities of taking contract-free jobs in the network largely affect an individual migrant's propensity. We provide three possible explanations on how networks influence migrants' contract decisions: job referral mechanism, limited information on contract benefits, and the "mini labor union" formed among co-villagers, which substitutes for a formal contract. In the empirical analysis, we apply the spatial autoregressive (SAR) linear probability model and the SAR logit model. By employing a comprehensive nationwide survey in 2011, we confirm that the common behavior of not signing contracts in the co-villager network increases the probability that a migrant accepts a contract-free job. The effects are larger for migrants whose jobs were introduced by their co-villagers, male migrants, migrants with rural Hukou, short-term migrants, and less educated migrants.

Keywords: *Contract, Co-Villager Network, Spatial Autoregressive Model, Internal Migrants*

JEL Classifications: *O15; R12; J41*

*Badi H. Baltagi: Department of Economics and Center for Policy Research, Syracuse University, 426 Eggers Hall, Syracuse, NY, 13244-1020, bbaltagi@maxwell.syr.edu; Ying Deng: University of International Business and Economics, 10 Huixin East Street, Chaoyang District, Beijing, China, 100029, ydeng.econ@gmail.com; Xiangjun Ma: University of International Business and Economics, xm2e@virginia.edu. All errors are the responsibility of the authors.

1 Introduction

Why would people accept a job without the protection of a labor contract? This is a decision made by around 40 percent of internal migrants in China based on a nationwide survey in 2011.[1] Generally speaking, an "irregular" job without a contract is not attractive due to lower wages, longer work hours, less insurance, *and so forth*. According to our data, these "irregular" jobs pay 22 percent lower hourly wages than the "regular" jobs that offer contracts at the mean, but demand 10 percent longer weekly work hours *on average*. Among the migrants with "regular" jobs, 35 percent have work-related insurance and 51 percent receive on-the-job training, while the figures are 3.75 percent and 21.4 percent for migrants with "irregular" jobs (see Table 1B). The welfare and return of workers are important for China to move from an economic pattern heavily dependent on exports to one in which the demand of the country's own internal market plays a larger role in economic growth (Becker and Elfstrom, 2010).

In addition to the average low education level of internal migrants, which leads to less competitive human capital and prevents them from finding better "regular" jobs, another important factor that affects their acceptance of "irregular" jobs is the co-villager network.[2] This paper shows that if most co-villagers (from the same home province residing in the same host city), who form the social network of an individual migrant, work for jobs without labor contracts, it is more likely that the individual migrant will also accept a job without a contract. Previous literature has documented the role that social networks play in migrants' work decisions, welfare participation, and migrating destinations (Calvo-Armengol and Jackson, 2004; Montgomery, 1991; Carrington et al., 1996). Studies on "Guanxi" (Chinese expression of social networks) in China have also received much attention (Zhang and Li, 2003; Lovett, Simmons and Kali, 1999). Chen, Jin and Yue (2010) find that one's migration decision is influenced by her co-villagers because co-villagers help each other in moving cost and job search at the destination.

From the perspective of migrant employees, incapability of finding better jobs, unawareness of full benefits bundled in a contract as well as job flexibility and bargaining power are the reasons we emphasize in this paper why internal migrants' decisions of accepting "irregular" jobs are affected by their co-villager networks.

First of all, many migrants find their jobs through co-villager networks. Calvo-Armengol and Jackson (2004) build an explicit network model on the transmission of job opening information, which leads to positive correlation between the employment status of agents who are directly or indirectly connected in the network of relatives, friends or acquaintances. Therefore, a contagion effect or network externality of employment status is observed. Ioannides and Loury (2004) conduct a comprehensive review on the effects of networks on job search.[3] Recently, Cingano and

[1] Although the Contract Law enacted on January 1, 2008 in China requires employers to provide written labor contracts for any position they offer, the law is not strictly implemented, especially during the global financial crisis right after the law was first launched. Becker and Elfstrom (2010) cited from the Beijing Federation of Trade Unions that only 32.8 percent of migrant workers had signed contracts by September 2008. They point out that the law specified the contract coverage but did not monitor the process of signing contracts well. In practice, costs of employers certainly rise if they obey the requests on wage and hour provisions as well as injury and social insurance. To avoid it, employers prefer to hire people who do not require contracts.

[2] In our sample, 66 percent of the migrants have an education level of middle school or below.

[3] For example, there is a positive correlation between getting assistance from a fraternity or sorority contacts and obtaining prestigious high-paying jobs (Marmaros and Sacerdote, 2002). Bayer, Ross and Topa (2008) find that a network, defined by residing in the same block, increases the probability of working together by over 33 percent.

Rosolia (2012) show that the increase in the employment rate of the network of former fellow workers reduces unemployment duration of individuals. However, the above research does not specify the co-villager network in China. In our data, more than 30 percent jobs of migrants are obtained through co-villager networks, while the corresponding figure in the literature is around 50%, summarized by Patacchini and Zenou (2012). They also show that members of a particular ethnic group concentrate in specific jobs and when new employment opportunities become available at their workplace, they pass this information along to contacts of the same race and ethnic background.

In spite of the easier job-searching process, the quality of jobs introduced by contacts is not necessarily high. Elliott (1999) finds that the use of informal contacts results in significantly lower wages for less-educated workers. Battisti, Peri and Romiti (2016) also find that immigrants initially located in places with larger co-ethnic networks are more likely to be employed at first, but earn lower wages in the long run. Loury (2006) claims that jobs obtained through contacts are better than those found through formal methods only when the contact is a prior-generation relative. For our analysis, if most co-villagers of an internal migrant work in enterprises that do not offer contracts, it is more likely that this migrant will be introduced to similar "irregular" jobs, whereas the probability to enter an enterprise offering "regular" jobs with contracts might be higher if the migrant searches jobs independently.

Secondly, migrants living and socializing in a large network acquire most information and knowledge from their co-villagers with similar background. As a result of information blocking on contract benefits, they do not fully or clearly understand the benefits (such as insurance and training) and legal protection bundling in a labor contract. Hence they imitate the behavior of their co-villagers. Co-villagers' higher propensity of accepting contract-free jobs increases the propensity for this migrant to accept the same type of jobs. In Bertrand, Luttmer and Mullainathan (2000)'s work on the usage of welfare programs, they conclude that social networks affect an individual's behavior and preference through two important channels: information and norms. With a randomized experiment, Duflo and Saez (2003) analyze the positive effects of information and social interactions on employees' decisions on enrollment in the retirement plan.

Lastly, collective bargaining from co-villager networks could substitute for labor contracts with respect to employment protection. To understand this channel, we first ask "if a person could choose between jobs with and without contracts, what can she benefit from the latter?" The answer is flexibility. Aguirregabira and Alonso-Borrego (2014) find that after restrictions on the more flexible fixed-term or temporary contracts were removed in the Spanish labor market, permanent workers were replaced by temporary workers.[4]Booth, Francesconi and Frank (2002) point out that temporary contracts in Europe are an important component of labor market flexibility, especially in countries characterized by high levels of employment protection. Measuring job flexibility by the years a migrant has worked for her current employer, we find in our sample that the average flexibility for jobs with and without contracts are 3.46 and 2.88 years, which means the latter is more flexible than the former. In other words, migrant workers prefer to access flexibility at the cost of lower wage and less employment security. This is particularly true for short-term job

Other literature on this issue includes Hellerstein et al. (2008), Falcon (2007), Munshi (2003), Wheatley et al. (2004), Battu et al. (2011), etc.

[4]In fact, fixed-term or temporary contracts are both considered as "regular" jobs with contracts in our sample. Nevertheless, the contract-free "irregular" jobs in our case are even more flexible but offer less protection for the employees.

seekers.[5]

On the other hand, collective bargaining makes it possible that a migrant may not need to worry much about job security if she works together with many co-villagers in the same workplace. In fact, a large proportion of employees could threaten the employer with a strike or to quit all together, which could prevent the institution from functioning. In this sense, the co-villager network can be regarded as a “mini labor union” which raises the bargaining power of the migrants. The insecurity of a contract-free job is thus reduced and the willingness of accepting it increases.

This paper applies a spatial autoregressive (SAR) model to examine how migrant workers' acceptance of contract-free jobs is affected by their co-villager networks. The SAR model, also known as the Cliff and Ord (1973) model, has been widely used in empirical analysis of social networks.[6]Lin (2010) applies the SAR model to identify the peer effects in students' academic achievements. Baltagi and Yen (2014) allow spatial correlation among neighboring hospitals and estimate the effects of externalities generated by competition and knowledge spillovers on hospital treatment rates. In our case, the dependent variable of the SAR model is a binary variable indicating whether a migrant accepts an irregular job without contract or not.

We employ the Dynamic Monitoring Survey of Internal Migrants in 2011 from the National Population and Family Planning Commission of China. This is a nationwide survey with 128,000 individual migrants residing in 326 host cities in all 31 provinces in China, which includes detailed demographic information and work information. In particular, we know whether the job of the migrant offers a contract as well as the home province and the host city of the migrant. In the benchmark SAR linear and logit regressions, we find that the probability of a migrant's acceptance of a job without a contract increases with the co-villager migrants' acceptance of contract-free jobs. The effects become larger for migrants whose jobs were introduced by their co-villagers, male migrants, migrants with rural Hukou, short-term migrants, and less educated migrants.[7]

To our knowledge, this is the first paper that uses a nationwide data set to study the interplay of co-villager networks and the decisions of accepting a contract-free irregular job offer for internal migrants in China. The remainder of the paper is organized as follows. Section 2 presents the SAR linear probability model and the SAR logit model. Section 3 describes data sources and reports summary statistics. Section 4 discusses the empirical results and analyzes the network effects in different subsamples. Section 5 concludes.

2 Methodology

*I*n the introduction, we present the conceptual framework that emphasizes three channels to explain why migrant workers are more likely to accept irregular jobs without contracts if more people in their co-villager networks accept such jobs: job searching process, information blocking on contract benefit, and collective bargaining of the mini labor union formed by co-villager networks at work. To test the hypothesis that when the co-villager migrants' acceptance of jobs without

[5]Most internal migrants in China are from the rural area. Some of them still have to do farm work at certain seasons of the year.

[6]LeSage and Pace (2009) have a detailed introduction of the SAR model. Its applications are widespread from studying spatial interactions at the macro level (countries, cities, etc.) to investigating social interactions at the micro level (households, individuals, etc.).

[7]Hukou is a social identity status in China, categorized into two types: rural or urban.

contracts increases, the probability increases for the targeting migrant to accept a job without a contract, we assign an outcome dummy of job types equal to 1 if an individual has a contract-free job, and 0 otherwise in the spatial autoregressive (SAR) model. Note that there are only 1.4 percent migrants who report unemployed in our sample, as most migrants are from the rural area. If they do not have a job in the host city, they will return home to do farm work. Therefore, unemployment is not an explicit issue for internal migrants in China.

We then assume that each co-villager has identical influential weight on a given individual, which is the inverse of the number of co-villagers in the network. Other migrants who are not co-villagers have zero influential weight. Therefore, for a given individual, taking the weighted average of the outcome dummy for all migrants provides us a measure of the acceptance of jobs without contracts in the co-villager network. This can be achieved by including a spatial lagged dependent variable in the regression.

In the next subsections, we first consider a spatial autoregressive (SAR) linear probability model to estimate the benchmark network effects. Based on the binary structure of the contract variable, we also employ an SAR logit model to confirm the network effects.

2.1 The SAR Linear Probability Model

The SAR linear probability model is given by

$$Contract_i = \lambda \sum_{j=1}^{n} w_{ij} \cdot Contract_j + X_i\beta + u_i, \quad i = 1, 2, ..., n, \tag{1}$$

where $Contract_i$ is a binary variable of migrant i's choice on accepting an irregular job without contract ($Contract_i = 1$) versus a regular job offering contract ($Contract_i = 0$), which can be interpreted as the probability that a migrant accepts a contract-free job. X_i is a $1 \times k$ vector of exogenous characteristics of migrant i, including the log value of hourly wage, gender, (rural or urban) Hukou status, marriage status, age, years the migrant has stayed in the residential city as well as the fixed effects of the current residential city, industry, occupation, type of employer, and the product of home province and education. β is a $k \times 1$ vector, which represents the coefficients of these exogenous regressors. The spatial lagged dependent variable $\sum_{j=1}^{n} w_{ij} \cdot Contract_j$ is a weighted average of the decisions on whether or not to accept a contract-free job of all other migrants in the co-villager network of migrant i. We specify the spatial weight matrix based on migrant i's co-villager network, which is defined as people flowing out of the same home province and residing in the same host city as migrant i. Only people within the network will make an impact on each other, as they are more likely to share information and provide job protection for each other. It is assumed that each migrant is equally affected by all others in the co-villager network. For example, if there are 40 migrants from Zhejiang province working in Beijing, then for each migrant, her Beijing-Zhejiang network consists of 39 people and each person is assigned a weight of 1/39 so as to satisfy the row-normalized condition of the spatial weight matrix. As a result, the spatial lagged dependent variable is the weighted average of the contracting decision of these 39 migrants in the network with the same weight of 1/39. Other migrants in the sample are excluded from the network and assigned zero weights. The diagonal elements are also set to zero as migrants are not considered as co-villagers of themselves. Therefore, the spatial weight matrix W is an $n \times n$ predetermined, row-normalized, symmetric, block-diagonal, sparse matrix, with a typical element w_{ij}

represents the weight of migrant j in i's network, hence, $w_{ii} = 0$ and $\sum_{j=1}^{n} w_{ij} = 1$ for $i = 1, 2, ..., n$. The coefficient λ captures the network effects on the probability of accepting a contract-free job.

An endogeneity issue arises as the decisions of taking contract-free irregular jobs of individuals in the network are also influenced by the decision of the objective individual in a symmetric way. In other words, $\sum_{j=1}^{n} w_{ij} \cdot Contract_j$ is correlated with the error term u_i. To achieve consistent estimates, we follow the spatial two-stage least squares (S2SLS) estimates proposed by Kelejian and Prucha (1998). They suggest using all the exogenous variables to construct a set of instruments for the endogenous spatial lagged dependent variable. Define $X = (X_1^{'}, X_2^{'}, ..., X_n^{'})'$ as an $n \times k$ matrix of all exogenous regressors, and then the instrument set is (X, WX).[8]

2.2 The SAR Logit Model

An SAR logit model could better accommodate the binary nature of our dependent variable. However, the nonlinear transformation of a logit model adds complexity in the estimation procedure as the typical maximum likelihood estimation (MLE) often involves n integrals in the likelihood function which can be burdensome when the sample size is large. Several approaches have been proposed to produce consistent estimates for the SAR model with a limited dependent variable. McMillen (1992) suggests an expectation–maximization (EM) algorithm to estimate the coefficients of the spatial probit model. Pinkse and Slade (1998) provide conditions of employing the generalized methods of moments (GMM) estimation to the spatial probit model. LeSage (2000) proposes Bayesian simulation approaches for SAR models.[9] Yet the computation intensities of these estimators depend highly on the sample size n, since they require the inversion of an $n \times n$ matrix. In fact, our data set includes more than 50,000 individuals, which makes it difficult to apply some of the estimation approaches mentioned above, even with strong computational power.

Klier and McMillen (2008) propose a linearized GMM (LGMM) approach, which is specifically designed for large samples. The linearized logit version of the spatial GMM estimator reduces estimation to two steps - standard logit followed by two-stage least squares. We adopt their LGMM approach and *consider* the SAR logit model

$$Contract_i^* = \lambda \sum_{j=1}^{n} w_{ij} \cdot Contract_j^* + X_i\beta + u_i, \quad i = 1, 2, ..., n, \tag{2}$$

$$Contract_i = I(Contract_i^* > 0),$$

where $Contract_i^*$ is a latent continuous variable measuring the propensity of migrant i accepting a contract-free irregular job versus a regular job offering contracts. Migrant $i's$ propensity depends upon the spatially weighted average of propensities of other migrants in her network, which is expressed as $\sum_{j=1}^{n} w_{ij} \cdot Contract_j^*$, where w_{ij} is the $(i, j)th$ element of the spatial weight matrix W, defined in the same way as in the last section. In fact, $Contract_i^*$ is unobservable. Instead, we can only observe a binary variable $Contract_i$ which is an indicator function of $Contract_i^*$ taking the value of 1 when migrant i has accepted an irregular job without contract, and 0 otherwise. Coefficient λ captures the network effects on the propensity of accepting jobs without contracts.

[8] See Lee (2003) and Kelejian et al. (2004) for discussions on the properties of S2SLS estimation with different sets of instruments.

[9] See LeSage and Pace (2009) and Smirnov (2010) for detailed reviews of the estimation methods for the spatial discrete choice models.

$\lambda > 0$ implies that higher propensities of accepting jobs without contracts of migrants in the co-villager network increases the propensity of accepting a contract-free irregular job of migrant i.

2.3 The Linearized GMM Estimation

We estimate the SAR logit model using a linearized version of the generalized method of moments (LGMM) suggested by Klier and McMillen (2008). Consider the matrix form of Equation (2)

$$Contract^* = \lambda W Contract^* + X\beta + u, \tag{3}$$

$$Contract = I(Contract^* > 0),$$

where $Contract^*$, $Contract$, and u are the $n \times 1$ vectors of $Contract_i^*$, $Contract_i$, and u_i, respectively. X is an $n \times k$ matrix of the k exogenous regressors. The reduced form can be written as

$$Contract^* = (I - \lambda W)^{-1} X\beta + (I - \lambda W)^{-1} u. \tag{4}$$

The variance-covariance matrix is proportional to $\Sigma = [(I - \lambda W)'(I - \lambda W)]^{-1}$, which implies heteroskedasticity and autocorrelation when spatial dependence exists, i.e. $\lambda \neq 0$. Denote the ith diagonal element of Σ as σ_i^2, we take into account the heteroskedasticity by defining $X_i^* = \frac{X_i}{\sigma_i}$ and $X^{**} = (I - \lambda W)^{-1} X^*$ where X^* is an $n \times k$ matrix of X_i^*.

As in Pinkse and Slade(1998), the generalized logit residual can be represented by $\varepsilon_i = Contract_i - P_i$, where $P_i = \frac{exp(X_i^{**}\beta)}{1+exp(X_i^{**}\beta)}$. Define the gradient terms as $G_i = (G_{\beta i}, G_{\lambda i})$, where $G_{\beta i} = \frac{\partial P_i}{\partial \beta} = P_i(1-P_i)X_i^{**}$ and $G_{\lambda i} = \frac{\partial P_i}{\partial \lambda} = P_i(1-P_i)[H_i\beta - \frac{X_i^{**}\beta}{\sigma_i^2}\Lambda_{ii}]$. H_i is the ith row of matrix $H = (I - \lambda W)^{-1} W X^{**}$, Λ_{ii} is the ith diagonal element of matrix $\Lambda = (I - \lambda W)^{-1} W (I - \lambda W)^{-1} (I - \lambda W)^{-1}$. When $\lambda = 0$, $(I - \lambda W)^{-1}$ degenerates to an identity matrix I. Thus, β can be consistently estimated by a standard logit model ignoring the spatial structure. The notation can also be greatly simplified as $X_i^{**} = X_i$. Let $\Gamma = (\beta', \lambda)'$ and $\Gamma_0 = (\hat{\beta}_0', 0)'$, where $\hat{\beta}_0$ is the estimate of β in the standard logit model. The gradient terms reduce to $G_{\beta i} = P_i^0(1-P_i^0)X_i$ and $G_{\lambda i} = P_i^0(1-P_i^0)H_i^0\beta_0$ when $\lambda = 0$, where H_i^0 is the ith row of an $n \times k$ matrix $H^0 = WX$.

Linearizing ε_i around the initial estimates of parameter Γ_0, we have

$$\varepsilon_i \approx \varepsilon_i^0 - G_i(\Gamma - \Gamma_0), \tag{5}$$

where $\varepsilon_i^0 = Contract_i - P_i^0$ and $P_i^0 = \frac{exp(X_i\hat{\beta}_0)}{1+exp(X_i\hat{\beta}_0)}$. Re-organizing equation (5) we can obtain

$$\varepsilon_i^0 + G_i\Gamma_0 \approx G_i\Gamma + \varepsilon_i. \tag{6}$$

Define Z_i as the ith row of a matrix of instruments Z. Consider a theoretical moment condition

$E(Z_i'\varepsilon_i) = 0$, the corresponding sample moment is thus

$$m(\beta,\lambda) = \frac{1}{n}\sum_{i=1}^{n} Z_i'\varepsilon_i. \tag{7}$$

A GMM estimator can be achieved by minimizing $\varepsilon' ZMZ'\varepsilon$ with respect to β and λ, where M is a positive definite weight matrix and ε is defined as the $n\times 1$ vectors of ε_i. Note that the computation load depends on the sample size n as it involves the inversion of an $n\times n$ matrix. Thus, the GMM estimator becomes infeasible when n is large. However, if M is set to $(Z'Z)^{-1}$, minimizing $\varepsilon' ZMZ'\varepsilon$ is equivalent as conducting a two-stage least squares (2SLS) estimation of a regression with ε as the error term and Z as the set of instruments. If applying on Equation (6), instead of minimizing $\varepsilon' Z(Z'Z)^{-1}Z'\varepsilon$ with respect to β and λ, the GMM estimator can be achieved by performing a 2SLS estimation of $\varepsilon_i^0 + G_i\Gamma_0$ with respect to G_i, employing a matrix of instruments Z.

In sum, the LGMM estimation procedure can be conducted in the following two steps:

Step 1: Estimate a standard logit model of *Contract* with respect to all the exogenous variables X to obtain a consistent estimate of β_0, $\hat{\beta}_0$. Then calculate the residuals ε_i^0 as well as the gradient terms $G_i = (G_{\beta i}, G_{\lambda i})$.

Step 2: Denote $\varepsilon^0 + G_\beta\hat{\beta}_0$, G_β and G_λ as the matrix counterparts of $\varepsilon_i^0 + G_{\beta i}\hat{\beta}_0$, $G_{\beta i}$, and $G_{\lambda i}$. Conduct a 2SLS estimation of $\varepsilon^0 + G_\beta\hat{\beta}_0$ with respect to G_β and G_λ, in which both G_β and G_λ are considered as endogenous variables and Z is a set of exogenous instruments chosen to deal with the endogeneity problem. More specifically, the 2SLS estimation involves the following two stage regressions:

- Stage 1: Regress G_β and G_λ on Z, respectively, to obtain the predicted values $\hat{G}_\beta$ and $\hat{G}_\lambda$.
- Stage 2: Regress $\varepsilon^0 + G_\beta\hat{\beta}_0$ on $\hat{G}_\beta$ and $\hat{G}_\lambda$. Thus, the corresponding coefficients of $\hat{G}_\beta$ and $\hat{G}_\lambda$ are the estimates of β and λ, respectively. In the empirical analysis, we employ $Z = (X, WX)$ as the instrument set for the LGMM estimation.

The advantage of the LGMM method is that no matrix needs to be inverted, because it requires only the standard logit and linear 2SLS estimations. The linearization significantly reduces the computation time and load as long as λ is small and the true structure is given by Equation (2). See Klier and McMillen (2008) for a detailed discussion of the finite sample properties of LGMM estimation.

3 Data and Summary Statistics

We describe the variables and data in this section. The main data source is the Dynamic Monitoring Survey of Internal Migrants in 2011 from the National Population and Family Planning Commission of China.[10] This is a nationwide survey with 128,000 individual migrants (who are 16-59 years old) residing in 326 host cities in 31 provinces all over the country. This survey includes individual demographic information, work information, and family information such as

[10]The survey has been conducted since 2010. We use the cross-sectional data in 2011 because the information on contracts is only available for this year.

age, gender, education, rural or urban Hukou status, marriage status, family members, current residential city, home province, wage, insurance, on-the-job training, industry, occupation, type of employer, how many years the migrant has stayed in the residential city, etc.[11] In particular, we know whether or not the migrant has signed a contract for the current job and whether their jobs were introduced by their relatives, classmates, or friends from hometown. After dropping the observations for whom contract information is missing as well as the outliers with the largest and smallest 1 percent of wage, we end up with 53,214 observations and 323 host cities in the sample.[12] These internal migrants come from 31 home provinces in China.

We first define a co-villager network as people from the same province and currently residing in the same host city. In spite of the large population of provinces in China, the average size of the network restricted in the host city is not very big. In our sample with 53,214 individuals, the median size of networks is 105 co-villagers. The mean and standard deviation of the network size are 232 and 304, separately.

We then define the dummy variable $Contract_i$ equal to 1 if the individual migrant has a job without contract, and 0 if the individual has a job with a contract, either fixed-term or non-fixed-term. In the full sample, 44.6 percent of the internal migrants have signed fixed-term contracts, 15.2 percent have signed non-fixed-term contracts, and 40.2 percent do not have any labor contract. The first column of Table 1A reports the percentage of the population who hold contracts, while the second column reports the percentage of those who do not hold contracts in the different subsamples. The last column presents the percentage of the type of migrants in the full sample. As shown, women and long-term migrants (who have stayed in the host cities for more than 1 year) are a little more likely to take regular jobs with contracts. More than 3/4 of the migrants with urban Hukou hold contracts, while those with rural Hukou are relatively more likely to accept contract-free jobs. Interestingly, people who found their job through networks, which means their jobs were introduced by their relatives, classmates, or friends from their hometown, are more likely to accept contract-free jobs than others. Moreover, the less educated the person is, the less likely she will hold a contract. A larger percentage of very old or very young migrants do not hold contracts. Lastly, married migrants have a lower probability of holding contracts.

Table 1B presents the summary statistics for the job-related variables in the samples with or without contracts. For the contract-free jobs, the mean of hourly wage is 9.4, 22 percent lower than the mean for the jobs with contracts (12.1), while the mean of weekly work hours are 10 percent higher (58.6 versus 53.3). The percentage of the population with insurance (including pension, medical insurance, injury insurance, unemployment insurance, maternity insurance, and housing funding) is significantly larger in the sample of regular jobs. The average percentage is 34.8, compared with 3.75 in the sample of contract-free jobs. In addition, only 21.4 percent of the workers with contract-free jobs have received skill training, in comparison with the 51.1 percent

[11] There are 15 industries as categorized in the survey including manufacture, mining, agriculture, forestry/pasture/husbandry/fishing, construction, electric/coal/water supply, wholesale and retail, lodging and catering, social service, finance/insurance/real estate, transportation/storage/communication, health care/sport/public welfare, education/culture/radio/movie/television, R&D/technology service, government/political organizations/social groups, and other industries; 12 types of employers including government or public institution, state owned enterprise, private business, Hong Kong, Macao, or Taiwan enterprise, Japanese or Korean enterprise, European or American enterprise, sino-foreign joint venture, and other enterprises; and 18 types of occupations including technical professionals, business related, etc.

[12] Most missing observations correspond to the self-employed migrants.

of contract-holders. These figures confirm that job openings that provide contracts are much better in terms of payments, weekly work hours, insurance, and training to employees.

Table 1C reports the summary statistics for job flexibility measured by the years a migrant has worked for her current employer in the samples with or without contracts separately. Job flexibility is reported as 0 when a migrant has worked for her current employer for less than 12 months. Apparently, migrants with contract-free irregular jobs hold their current jobs for a shorter period on average. In other words, the mobility, which can be considered as the revealed outcome of flexibility of this type of jobs is higher. The mean, the percentiles as well as the standard deviation of these jobs are all smaller than the corresponding statistics for the regular jobs that provide contracts.

Table 1D demonstrates the percentage of migrants with different work experience in the current job in the contract or non-contract samples. The cutoffs correspond to the minimum value, 25 percentile, median, 75 percentile and 99 percentile of job flexibility in the full sample. For example, there are 17.78 percent and 26.08 percent of migrants who have worked for her current employer for less than 12 months in the samples with or without contracts. The percentage in the non-contract sample is larger than that in the contract sample when the period a migrant has worked for her current employer is less than or equal to 1 year. However, when the work period is longer than 2 years (the median in the full sample), 4 years (the 75 percentile in the full sample) or 20 years (the 99 percentile in the full sample), the opposite case shows up. That is, relatively more migrants with contract-free irregular jobs stick to their current jobs for a shorter time whereas relatively more migrants with regular jobs that offer contracts hold their jobs for a longer time, which indicates the revealed flexibility of the former type of jobs.

4 Empirical Results and Subsample Analysis

4.1 Basic Results of the Full Sample

Table 2 presents the basic estimation results of the SAR linear probability and the SAR logit models, respectively. As specified in Section 2, we use $Z = (X, WX)$ as a set of instruments to deal with the endogeneity problem resulting from the spatial lagged dependent variable. X represents a set of exogenous variables. Specifically, Columns (1) and (4) only control for individual demographic variables including the log value of hourly wage, gender (a dummy equal to 1 if gender is male, and 0 otherwise), Hukou status (a dummy equal to 1 if urban, and 0 if rural), marriage status (a dummy equal to 1 if married or have married, and 0 if single), age, years the migrant has stayed in the residential city. In Columns (2) and (5) we add the fixed effects of the current residential city, industry, occupation, type of employer. Columns (3) and (6) also control for the product of home province and education so as to control for the unobserved heterogeneous factors at the home-province-education level.

Both the SAR linear probability model (using the S2SLS estimation method) and the SAR logit model (using the LGMM estimation method) show significant positive coefficient estimates of the spatial lagged dependent term. This implies that a migrant's probability of accepting a job that does not provide a contract increases with the decisions of other migrants in her co-villager network who accept jobs without contracts. However, the inclusion of more control variables lead to a large reduction of the network effects, dropping from 0.5274 to 0.0975 for the linear probability model and from 0.7265 to 0.0467 for the logit model. Taking Column (3) as the baseline outputs for the SAR linear probability model with the most strict controls, the result suggests that when the percentage of migrants in the co-villager network accepting jobs without contracts increases by 10%, the probability of the targeting migrant accepting a job without a contract increases by about 1%. We understand that besides network, there might be other possibilities that may explain the results we find. For instance, it may, to some extent, represent the magnitude of information or skills collected by people migrating from the same province to the same host city. Therefore, we caution that the estimation we obtained could be biased up.

In the meantime, our estimates imply that higher wage is associated with lower probability of taking irregular jobs that does not offer contract. Male migrants and migrants with rural Hukou are more likely to accept contract-free irregular jobs, presumably because there are more irregular job opportunities for these types of migrants. As we observe in the statistics, married migrants and long-term migrants are also more likely to accept irregular jobs. For married migrants, one possible explanation is that they have families back in their home province, and thus they value flexibility more as they need to visit their families more often. For migrants who have stayed longer in the residential city, they might have entered the informal labor market without contracts years ago, at which time there was no law to regulate contract implementation. They may have been blocked (on information) and got stuck (due to work experience and job qualification) in the informal labor market since then. Lastly, the linear probability model also shows that younger migrants are more likely to accept the contract-free jobs. This should be tempered by the fact that this is not significant in the logit model.

4.2 Migrants Who Found Jobs through Networks

Next, we identify the more effective co-villager networks by dividing the sample into two subsets: (1) migrants who found their jobs through the co-villager networks, that is, their jobs were introduced by their relatives, classmates, or friends from hometown; (2) migrants who found their jobs through other channels, such as local friends in the host city, the internet, job-searching agents, and so on. We refer the former case as the group of migrants with more effective co-villager networks. We anticipate that migrants who found their jobs through their co-villager networks tend to work together with their co-villagers. Thus the function of employment protection of networks is more likely to take into effect. Therefore, we expect a larger network effect in the first subsample.

We utilize the same network weight matrix constructed in the full sample and run regressions in both subsamples, respectively. Table 3 only reports the estimates containing the same controls as in Columns (3) and (6) in Table 2 using the S2SLS and LGMM methods. As predicted, the network effects in Subsample (1) are much larger than those in Subsample (2) regardless of the estimation methods. Taking Columns (1) and (2) as the baseline outputs obtained from the SAR models, we find that the network effect in the subsample who found their jobs through co-villagers is more than twice of the network effect in the subsample whose jobs were not introduced by co-villagers.

4.3 Robustness Check for Other Subsamples

Further, we conduct analysis in the following subsamples: (1) males versus females; (2) migrants with rural Hukou versus urban Hukou; (3) long-term migration (people who have stayed in the residential city for more than a year) versus short-term migration (people who have stayed in the residential city for a year or less than a year). We still use the co-villager network weight matrix constructed in the full sample, but run regressions in these subsamples. Tables 4-6 only report estimates containing the most strict controls as in Columns (3) and (6) in Table 2 using the S2SLS and LGMM methods, respectively. These findings show that the decisions of male migrants, migrants with rural Hukou, and short-term migrants are more likely to be influenced by networks. They may rely more on co-villager networks to find jobs, be less aware of the benefits offered by contracts, or depend more on the co-villager colleagues to bargain with their employers. In fact, network effects become insignificant for women, migrants with urban Hukou, and long-term migrants in the SAR logit model.

4.4 Network Effects for Migrants of Different Education Levels

In Table 7, we investigate the network effects for migrants of different education levels: primary school or below, middle school, high school or equivalent, and college or equivalent or above. Note that we include fixed effects of the current residential province instead of the current residential city as there are fewer observations in some subsamples. Fixed effects of education and the product of education and home province from regressions are also excluded. The SAR logit model estimates reveal that significant network effects only exist among migrants with education levels below college. Although the SAR linear probability model estimate of network effects for migrants with college education or above is significant, the magnitude is much lower than those in other subsamples. To conclude, migrants with lower education levels rely more on co-villager networks.

4.5 Network Effects in Different Industries

We also investigate network effects in different industries. The top 4 industries in the survey are manufacturing, construction, lodging and catering as well as wholesale and retail. Table 8 reports the S2SLS and LGMM estimates of these industries. We use the same control variables, i.e., the same individual demographic variables and fixed effects as in the basic model (Table 2), but excluding industry fixed effects and the product of home province and education. Again, host city fixed effects are replaced with host province fixed effects due to fewer observations. As we can see, the network has the largest effect in the manufacturing industry, followed by construction and lodging and catering industries. This shows that migrants working in these industries are more likely to find jobs through networks or rely on networks to gain benefits. For example, a construction project in China is often contracted out to a team in which the members are usually recruited from the same village. Therefore, the migration decision, job search process, and thus the contract decision of these people are strongly influenced by the co-villager network. Moreover, it is easier for these co-villagers who work in the same workplace to form the "mini labor union," which could take the place of a formal contract.

Interestingly, the network effect is insignificant in the wholesale and retail industry. This may arise from the smaller sample size as many migrants in this industry are self-employed and there is no contract information in the data.

4.6 Network Effects in Different Home Provinces

Lastly, Table 9 demonstrates the heterogeneous network effects for migrants from different home provinces. Sichuan, Henan, Anhui, and Shangdong are the 4 largest provinces from which migrants move out. The proportion of migrants in our data from these provinces are 10.32%, 8.24%, 7.58%, and 5.80%. We run regressions in these 4 subsamples with the host province, industry, occupation, type of employer and education fixed effects. The SAR linear probability model shows significant network effects in all four provinces, and the effects are stronger in Sichuan and Shandong but relatively weaker in Henan and Anhui. The SAR logit model finds insignificant network effect in Anhui and the strongest effect in Shandong. Different culture and norms across home provinces may explain the various network effects estimated.

5 Conclusion

This paper employs a comprehensive data set from a nationwide survey to study how co-villager networks affect internal migrants' decision of accepting irregular jobs without contracts in the labor market of China. Surrounded by a network with a large amount of co-villagers work for business that do not provide contracts, the migrant not only has limited job choices, but also knows little about the benefits a regular labor contract would offer because of the lack of information. Additionally, the network serves as a "mini labor union" that raises the bargaining power of the migrants and substitutes for the contract with respect to employment protection. As a result, the migrant tends to accept a similar contract-free job offer that features flexibility. We include a spatial lagged dependent term to capture the network effect. Besides the SAR linear probability model estimated by the S2SLS method, a linearized GMM approach is conducted to estimate the SAR logit model given the large data set. Our empirical work confirms that the probability of a migrant's acceptance of jobs without contracts increases with the acceptance of jobs without contracts of other migrants in her co-villager network. We further find that the network effects are larger in the subsamples of migrants whose jobs were introduced by their co-villagers, male migrants, migrants with rural Hukou, short-term migrants, and less educated migrants. The network effects also vary by industries and home provinces.

Appendix

Table 1: Summary Statistics

Table 1A: Percentage of Population with and without Contracts

		Has Contract (%)	No Contract (%)	Percentage in the Full Sample (%)
Full Sample		59.83	40.17	-
Sex	*Male*	59.06	40.94	58.47
	Female	60.92	39.08	41.53
Hukou	*Rural*	56.69	43.31	82.89
	Urban	75.07	24.93	17.11
Migrating Years	*Equal to or Less than 1 Year*	58.06	41.94	39.21
	More than 1 Year	60.98	39.02	60.79
Found Job	*Yes*	57.54	42.46	37.75
through Network	*No*	61.22	38.78	62.25
Education	*School Dropouts*	39.72	60.28	1.33
	Primary School	43.44	56.56	12.12
	Middle School	54.93	45.07	52.13
	High School	66.14	33.86	15.55
	Technical Secondary School	73.60	26.40	7.40
	Junior College	80.96	19.04	7.57
	College	90.71	9.29	3.68
	Graduate School	94.74	5.26	0.21
Age	*[16, 18)*	44.43	55.57	1.40
	[18, 30)	63.24	36.76	45.47
	[30, 40)	60.40	39.60	31.22
	[40, 50)	53.59	46.41	18.93
	[50, 59]	48.83	51.17	2.98
Marriage	*Married*	58.83	41.17	67.67
	Single	61.94	38.06	32.33

Note: Table 1A reports the percentage of population who have contracts or not in different subsamples of (1) males or females, (2) rural or urban Hukou, (3) long-term migrants (staying in the host city for more than a year) or short-term migrants (staying in the host city for a year or less than a year), (4) migrants who found their jobs through networks or not, (5) different education levels, (6) different ages, and (7) different marriage status.

Table 1B: Statistics of Job-Related Variables in the Contract or Non-Contract Samples

		# of Obs.	Mean	Std. Dev.	Min.	Median	Max.
Hourly Wage	*Full Sample*	53214	11.05	6.03	2.98	9.38	43.75
	Has Contract	31840	12.14	6.47	2.98	10.42	43.75
	No Contract	21374	9.43	4.87	2.98	8.33	43.75
Weekly Working Hours	*Full Sample*	53214	55.45	13.29	2	54	112
	Has Contract	31840	53.32	12.69	4	48	112
	No Contract	21374	58.62	13.53	2	56	112

		Full Sample		**Has Contract**		**No Contract**	
		Yes (%)	**No (%)**	**Yes (%)**	**No (%)**	**Yes (%)**	**No (%)**
Insurance	*Pension*	28.29	71.71	44.63	55.37	3.94	96.06
	Medical Insurance	31.90	68.10	48.84	51.16	6.66	93.34
	Injury Insurance	33.66	66.34	50.43	49.57	8.67	91.33
	Unemployment Insurance	18.56	81.44	29.95	70.05	1.60	98.40
	Maternity Insurance	13.24	86.76	21.39	78.61	1.11	98.89
	Housing Fund	8.44	91.56	13.76	86.24	0.51	99.49
	Average	22.35	77.65	34.83	65.17	3.75	96.25
Work Skill Training		39.15	60.85	51.10	48.90	21.36	78.64

Note: Table 1B presents the summary statistics for the job-related variables in samples with or without contracts, separately. The job-related variables include hourly wages, insurance (i.e. pension, medical insurance, injury insurance, unemployment insurance, maternity insurance, and housing funding) as well as work skill training.

Table 1C: Statistics of Job Flexibility in the Contract or Non-Contract Samples

Years a Migrant Has Worked for the Current Employer		# of Obs.	Mean	Std. Dev.
Job Flexibility	*Full Sample*	53153	3.22	4.17
	Has Contract	31808	3.46	4.31
	No Contract	21374	2.88	3.95

Note: Table 1C presents the summary statistics for job flexibility measured by the years a migrant has worked for her current employer in samples with or without contracts, separately. Job flexibility is reported as 0 when a migrant has worked for her current employer for less than 12 months.

Table 1D: Percentage of Migrants with Different Work Experience in the Current Employer in the Contract or Non-Contract Samples

Years a Migrant Has Worked for the Current Employer		Has Contract (%)	No Contract (%)
Job Flexibility	*Less than 1 year (*job flexibility=*0)*	17.78	26.08
	Less than or equal to 1 year (25% of job flexibility in the full sample)	41.89	51.02
	More than 2 years (50% of job flexibility in the full sample)	42.31	34.33
	More than or equal to 4 years (75% of job flexibility in the full sample)	32.28	25.79
	More than or equal to 20 years (99% of job flexibility in the full sample)	1.31	1.00

Note: Table 1D presents the percentage of migrants with different work experience in the current job in the contract or non-contract samples. For example, there are 17.78 percent and 26.08 percent of migrants who have worked for her current employer for less than 12 months in the samples with or without contracts, separately. The cutoffs are the minimum value, 25 percentile, median, 75 percentile and 99 percentile of job flexibility in the full sample.

Table 2: Basic Regressions in the Full Sample

	Dependent variable: *Contract*					
	SAR Model			SAR Logit Model		
	(1)	(2)	(3)	(4)	(5)	(6)
Network Contract	0.5274***	0.1097***	0.0975***	0.7265***	0.1546***	0.0467**
	(0.0173)	(0.0215)	(0.0226)	(0.0286)	(0.0360)	(0.0186)
ln(Wage)	-0.2073***	-0.1159***	-0.1167***	-1.4167***	-0.7364***	-0.9900***
	(0.0046)	(0.0047)	(0.0049)	(0.0478)	(0.0582)	(0.0511)
Sex	0.0489***	0.0147***	0.0170***	0.3829***	0.0893***	0.1474***
	(0.0042)	(0.0042)	(0.0044)	(0.0224)	(0.0257)	(0.0281)
Hukou Status	0.1068***	0.0308***	0.0313***	0.9085***	0.2630***	0.4813***
	(0.0054)	(0.0058)	(0.0061)	(0.0318)	(0.0513)	(0.0522)
Marital Status	0.0192***	0.0126**	0.0109*	0.1432***	0.0600*	0.0638*
	(0.0055)	(0.0053)	(0.0055)	(0.0273)	(0.0319)	(0.0351)
Age	0.0022***	-0.0007**	-0.0008**	0.0132***	-0.0031*	-0.0015
	(0.0003)	(0.0003)	(0.0003)	(0.0013)	(0.0016)	(0.0018)
Stay Years	-0.0005	0.0012***	0.0015***	-0.0024	0.0067**	0.0055*
	(0.0005)	(0.0004)	(0.0005)	(0.0022)	(0.0026)	(0.0029)
Fixed effects						
Host City	No	Yes	Yes	No	Yes	Yes
Home Province	No	Yes	Yes	No	Yes	Yes
Industry	No	Yes	Yes	No	Yes	Yes
Occupation	No	Yes	Yes	No	Yes	Yes
Type of Employer	No	Yes	Yes	No	Yes	Yes
Education	No	Yes	Yes	No	Yes	Yes
Home Province*Education	No	No	Yes	No	No	Yes
# of Observations	53214	53214	53214	53214	53214	53214
R-Squared	0.1461	0.2788	0.2249	0.0378	0.1044	0.2130

Note: Columns (1), (2), and (3) are the spatial 2SLS estimates with different sets of controls, while Columns (4), (5), and (6) are the LGMM estimates with the corresponding sets of controls. Standard errors are in parenthesis. ***, **, * indicate significance at the 1%, 5% and 10% level.

Table 3: Regressions in Subsamples of Migrants Who Found Jobs through Co-Villagers or Not

	Dependent variable: *Contract*			
	SAR Model		SAR Logit Model	
	Job Was Introduced by Co-Villagers	Job Was Not Introduced by Co-Villagers	Job Was Introduced by Co-Villagers	Job Was Not Introduced by Co-Villagers
	(1)	(2)	(3)	(4)
Network Contract	0.1806***	0.0751***	0.0744**	0.0540**
	(0.0355)	(0.0275)	(0.0301)	(0.0247)
ln(Wage)	-0.0939***	-0.1247***	-0.5362***	-1.0143***
	(0.0083)	(0.0058)	(0.0520)	(0.0688)
Sex	0.0262***	0.0093*	0.1480***	0.0682**
	(0.0071)	(0.0053)	(0.0420)	(0.0342)
Hukou Status	0.0255**	0.0269***	0.1876***	0.3840***
	(0.0110)	(0.0069)	(0.0708)	(0.0601)
Marital Status	0.0132	0.0102	0.0675	0.0209
	(0.0091)	(0.0066)	(0.0515)	(0.0441)
Age	-0.0012**	-0.0007**	-0.0060**	-0.0030
	(0.0005)	(0.0004)	(0.0026)	(0.0023)
Stay Years	0.0009	0.0012**	0.0060	0.0064*
	(0.0008)	(0.0006)	(0.0045)	(0.0035)
Fixed effects	Host City, Home Province, Industry, Occupation, Type of Employer, Education, Home Province*Education			
# of Observations	20088	33126	20088	33126
R-Squared	0.2732	0.2999	0.2519	0.2912

Note: Columns (1) and (3) report the estimates in the subsample of migrants who found jobs through co-villager networks, while Columns (2) and (4) report the estimates for those who did not find job through co-villagers. Columns (1) and (2) are the spatial 2SLS estimates while Columns (3) and (4) are the LGMM estimates. In all regressions, we control for the fixed effects of the host city, home province, industry, occupation, type of employer, education, and the interactions of home province and education. Standard errors are in parentheses. ***, **, and * indicate significance at the 1%, 5% and 10% level.

Table 4: Regressions in Subsamples of Male/Female Migrants

	Dependent variable: *Contract*			
	SAR Model		SAR Logit Model	
	Males	Females	Males	Females
	(1)	(2)	(3)	(4)
Network Contract	0.1247***	0.0822**	0.0678***	0.0041
	(0.0263)	(0.0384)	(0.0210)	(0.0273)
ln(Wage)	-0.1029***	-0.1377***	-0.6444***	-0.9168***
	(0.0062)	(0.0077)	(0.0470)	(0.0669)
Hukou Status	0.0313***	0.0263***	0.2644***	0.2630***
	(0.0077)	(0.0091)	(0.0577)	(0.0640)
Marital Status	0.0122*	0.0079	0.0647	0.0299
	(0.0071)	(0.0084)	(0.0426)	(0.0528)
Age	-0.0010***	-0.0003	-0.0052**	-0.0017
	(0.0004)	(0.0005)	(0.0020)	(0.0030)
Stay Years	0.0019***	-0.0000	0.0122***	-0.0020
	(0.0006)	(0.0008)	(0.0034)	(0.0046)
Fixed effects	Host City, Home Province, Industry, Occupation, Type of Employer, Education, Home Province*Education			
# of Observations	31115	22099	31115	22099
R-Squared	0.2917	0.2908	0.3665	0.3801

Note: Columns (1) and (3) report the estimates in the subsample of male migrants, while Columns (2) and (4) report the estimates of the female subsample. Columns (1) and (2) are the spatial 2SLS estimates, while Columns (3) and (4) are the LGMM estimates. In all regressions, we control for the fixed effects of the host city, home province, industry, occupation, type of employer, education, and the interactions of home province and education. Standard errors are in parentheses. ***, **, and * indicate significance at the 1%, 5% and 10% level.

Table 5: Regressions in Subsamples of Migrants with Rural/Urban Hukou

	Dependent variable: *Contract*			
	SAR Model		SAR Logit Model	
	Rural Hukou	Urban Hukou	Rural Hukou	Urban Hukou
	(1)	(2)	(3)	(4)
Network Contract	0.1034***	0.1382***	0.0832***	0.0243
	(0.0251)	(0.0417)	(0.0253)	(0.0468)
ln(Wage)	-0.1180***	-0.1088***	-0.6798***	-0.8861***
	(0.0054)	(0.0096)	(0.0374)	(0.1346)
Sex	0.0155***	0.0139	0.0861***	0.1032
	(0.0048)	(0.0088)	(0.0271)	(0.0758)
Marital Status	0.0139**	-0.0034	0.0758**	-0.0660
	(0.0061)	(0.0108)	(0.0347)	(0.0961)
Age	-0.0009***	0.0004	-0.0048***	0.0059
	(0.0003)	(0.0006)	(0.0018)	(0.0051)
Stay Years	0.0009*	0.0023**	0.0054*	0.0171**
	(0.0005)	(0.0010)	(0.0029)	(0.0079)
Fixed effects	Host City, Home Province, Industry, Occupation, Type of Employer, Education, Home Province*Education			
# of Observations	44107	9107	44107	9107
R-Squared	0.2677	0.3403	0.2150	0.3428

Note: Columns (1) and (3) report the estimates in the subsample of migrants with rural Hukou, while Columns (2) and (4) report the estimates of subsample of migrants with urban Hukou. Columns (1) and (2) are the spatial 2SLS estimates, while Columns (3) and (4) are LGMM estimates. In all regressions, we control for the fixed effects of the host city, home province, industry, occupation, type of employer, education, and the interactions of home province and education. Standard errors are in parentheses. ***, **, and * indicate significance at the 1%, 5% and 10% level.

Table 6: Regressions in Subsamples of Short-Term/Long-Term Migrants

	Dependent variable: *Contract*			
	SAR Model		SAR Logit Model	
	Long-Term Migrants	Short-Term Migrants	Long-Term Migrants	Short-Term Migrants
	(1)	(2)	(3)	(4)
Network Contract	0.1314***	0.0883***	0.0114	0.0483**
	(0.0289)	(0.0330)	(0.0211)	(0.0225)
ln(Wage)	-0.1198***	-0.1021***	-0.7643***	-0.5877***
	(0.0059)	(0.0081)	(0.0543)	(0.0507)
Sex	0.0137**	0.0200***	0.0831**	0.1076***
	(0.0054)	(0.0068)	(0.0338)	(0.0394)
Hukou Status	0.0299***	0.0342***	0.2516***	0.2565***
	(0.0071)	(0.0103)	(0.0548)	(0.0677)
Marital Status	0.0112	0.0181**	0.0611	0.1001**
	(0.0070)	(0.0085)	(0.0449)	(0.0489)
Age	-0.0002	-0.0013***	-0.0007	-0.0066**
	(0.0004)	(0.0005)	(0.0022)	(0.0027)
Fixed effects	Host City, Home Province, Industry, Occupation, Type of Employer, Education, Home Province*Education			
# of Observations	32347	20867	32347	20867
R-Squared	0.3062	0.2724	0.4535	0.3965

Note: Columns (1) and (3) report the estimates in the subsample of long-term migrants, while Columns (2) and (4) report the estimates of subsample of short-term migrants. Columns (1) and (2) are the spatial 2SLS estimates, while Columns (3) and (4) are the LGMM estimates. In all regressions, we control for the fixed effects of the host city, home province, industry, occupation, type of employer, education, and the interactions of home province and education. Standard errors are in parentheses. ***, **, and * indicate significance at the 1%, 5% and 10% level.

Table 7: Regressions in Subsamples of Migrants of Different Education Levels

	Dependent variable: *Contract*							
	SAR Model				SAR Logit Model			
	Primary School or under	Middle School	High School or Equivalent	College or Equivalent or above	Primary School or under	Middle School	High School or Equivalent	College or Equivalent or above
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Network Contract	*0.1865****	*0.2015****	*0.2001****	*0.0809***	*0.1693****	*0.1325****	*0.1237****	*-0.0210*
	(0.0450)	*(0.0244)*	*(0.0354)*	*(0.0360)*	*(0.0555)*	*(0.0285)*	*(0.0363)*	*(0.0739)*
ln(Wage)	*-0.0978****	*-0.1173****	*-0.1476****	*-0.0990****	*-0.4584****	*-0.5773****	*-0.7889****	*-1.0961****
	(0.0136)	*(0.0069)*	*(0.0092)*	*(0.0103)*	*(0.0847)*	*(0.0414)*	*(0.1530)*	*(0.2637)*
Sex	*-0.0123*	*0.0263****	*0.0247****	*0.0048*	*-0.0546*	*0.1254****	*0.1196***	*0.0669*
	(0.0126)	*(0.0062)*	*(0.0083)*	*(0.0092)*	*(0.0638)*	*(0.0335)*	*(0.0552)*	*(0.0930)*
Hukou Status	*0.1061****	*0.0194**	*0.0343****	*0.0451****	*0.5598****	*0.1051***	*0.2024****	*0.5005****
	(0.0304)	*(0.0100)*	*(0.0093)*	*(0.0096)*	*(0.1580)*	*(0.0531)*	*(0.0772)*	*(0.1444)*
Marital Status	*0.0331**	*0.0333****	*-0.0086*	*-0.0203**	*0.1565*	*0.1685****	*-0.0444*	*-0.2520**
	(0.0201)	*(0.0079)*	*(0.0101)*	*(0.0107)*	*(0.1026)*	*(0.0423)*	*(0.0635)*	*(0.1301)*
Age	*-0.0015***	*-0.0016****	*0.0012**	*0.0019***	*-0.0072**	*-0.0078****	*0.0061*	*0.0212***
	(0.0007)	*(0.0004)*	*(0.0006)*	*(0.0009)*	*(0.0039)*	*(0.0021)*	*(0.0043)*	*(0.0099)*
Stay Years	*0.0025***	*0.0018****	*-0.0010*	*0.0010*	*0.0130***	*0.0106****	*-0.0046*	*0.0045*
	(0.0010)	(0.0006)	(0.0010)	(0.0011)	(0.0055)	(0.0034)	(0.0069)	(0.0114)
Fixed effects	Host Province, Home Province, Industry, Occupation, Type of Employer							
# of Observations	7160	27741	12211	6102	7160	27741	12211	6102
R-Squared	0.2079	0.2265	0.2469	0.2106	0.2207	0.148	0.1826	0.1099

Note: Columns (1)-(4) report the spatial 2SLS estimates in the subsamples of primary school or below, middle school, high school or equivalent as well as college or above, respectively. Columns (5)-(8) are the LGMM estimates in the corresponding subsamples. In all regressions, we control for the fixed effects of the host province, home province, industry, occupation, and type of employer. Standard errors are in parentheses. ***, **, and * indicate significance at the 1%, 5% and 10% level.

Table 8: Regressions in Subsamples of Different Industries

	Dependent variable: *Contract*							
	SAR Model				SAR Logit Model			
	Manufacturing	Construction	Lodging & Catering	Wholesale & Retail	Manufacturing	Construction	Lodging & Catering	Wholesale & Retail
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Network Contract	0.2806***	0.2581***	0.2196***	0.0975	0.2329***	0.1123**	0.1252**	0.0311
	(0.0288)	(0.0420)	(0.0536)	(0.0613)	(0.0389)	(0.0527)	(0.0568)	(0.0747)
ln(Wage)	-0.1043***	-0.0732***	-0.1826***	-0.1791***	-0.4329***	-0.4437***	-0.9708***	-0.9576***
	(0.0081)	(0.0135)	(0.0162)	(0.0177)	(0.1411)	(0.0801)	(0.0948)	(0.1000)
Sex	0.0272***	-0.0106	0.0084	0.0506***	0.1099*	-0.1297	0.0382	0.2488***
	(0.0062)	(0.0172)	(0.0129)	(0.0159)	(0.0565)	(0.1096)	(0.0614)	(0.0852)
Hukou Status	0.0346***	0.0646***	0.0055	0.0276	0.1071	0.3470***	0.0196	0.2055*
	(0.0102)	(0.0199)	(0.0187)	(0.0205)	(0.1499)	(0.1192)	(0.0898)	(0.1230)
Marital Status	0.0283***	0.0357**	0.0174	0.0094	0.1303**	0.2061**	0.0923	0.0094
	(0.0083)	(0.0170)	(0.0176)	(0.0204)	(0.0641)	(0.0950)	(0.0840)	(0.1089)
Age	-0.0009**	-0.0016**	-0.0015	-0.0005	-0.0044	-0.0104**	-0.0078	0.0002
	(0.0005)	(0.0007)	(0.0010)	(0.0013)	(0.0030)	(0.0041)	(0.0050)	(0.0071)
Stay Years	-0.0018**	0.0073***	0.0004	0.0015	-0.0077	0.0602***	0.0042	0.0095
	(0.0007)	(0.0012)	(0.0016)	(0.0019)	(0.0050)	(0.0115)	(0.0081)	(0.0105)
Fixed effects	Host Province, Home Province, Occupation, Type of Employer, Education							
# of Observations	18850	6430	6008	3839	18850	6430	6008	3839
R-Squared	0.2295	0.2757	0.1555	0.2333	0.1670	0.2000	0.1026	0.1174

Note: Columns (1)-(4) report the spatial 2SLS estimates in the subsamples of migrants in manufacturing, construction, lodging and catering as well as wholesale and retail industries, respectively. Columns (5)-(8) are the LGMM estimates for the corresponding subsamples. In all regressions, we control for the fixed effects of the host province, home province, occupation, type of employer, education, and the interactions of home province and education. Standard errors are in parentheses. ***, **, and * indicate significance at the 1%, 5% and 10% level.

Table 9: Regressions in Subsamples of Migrants from Different Home Provinces

	Dependent variable: *Contract*							
	SAR Model				SAR Logit Model			
	Shandong	Henan	Sichuan	Anhui	Shandong	Henan	Sichuan	Anhui
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Network Contract	0.5139***	0.3610***	0.3061***	0.5193***	0.1362**	0.1890**	0.0959	0.3298***
	(0.0475)	(0.0738)	(0.0780)	(0.0668)	(0.0574)	(0.0785)	(0.1093)	(0.1110)
ln(Wage)	-0.1370***	-0.1167***	-0.1175***	-0.1465***	-0.8401***	-0.6177***	-0.5487***	-1.3693***
	(0.0147)	(0.0172)	(0.0182)	(0.0199)	(0.0978)	(0.1090)	(0.1296)	(0.2350)
Sex	0.0100	0.0198	0.0445***	0.0196	0.0361	0.1108	0.1975**	0.1719
	(0.0132)	(0.0144)	(0.0157)	(0.0167)	(0.0747)	(0.0848)	(0.0943)	(0.1120)
Hukou Status	0.0122	0.0439*	0.0626**	0.0392	0.2192*	0.2855	0.3012	0.6530**
	(0.0199)	(0.0263)	(0.0254)	(0.0262)	(0.1271)	(0.1747)	(0.2178)	(0.2600)
Marital Status	0.0310*	0.0152	-0.0185	0.0531**	0.1805*	0.0867	-0.0954	0.4533***
	(0.0177)	(0.0189)	(0.0210)	(0.0219)	(0.1003)	(0.1129)	(0.1272)	(0.1636)
Age	-0.0024***	-0.0017	0.0006	0.0005	-0.0133***	-0.0077	0.0020	-0.0018
	(0.0009)	(0.0011)	(0.0011)	(0.0013)	(0.0049)	(0.0062)	(0.0065)	(0.0076)
Stay Years	-0.0014	-0.0006	-0.0023	-0.0014	-0.0110	-0.0051	-0.0108	-0.0014
	(0.0013)	(0.0020)	(0.0018)	(0.0018)	(0.0077)	(0.0114)	(0.0103)	(0.0112)
Fixed effects	Host Province, Industry, Occupation, Type of Employer, Education							
# of Observations	5491	4384	4031	3084	5477	4384	4022	3065
R-Squared	0.2760	0.2834	0.2557	0.2477	0.3855	0.2660	0.1343	0.0779

Note: Columns (1)-(4) report the spatial 2SLS estimates in the subsamples of migrants from Shandong, Henan, Sichuan, and Anhui province, respectively. Columns (5)-(8) are the LGMM estimates for the corresponding subsamples. In all regressions, we control for the fixed effects of the host province, industry, occupation, type of employer, and education. Standard errors are in parentheses. ***, **, and * indicate significance at the 1%, 5% and 10% level.

References

[1] Aguirregabiria, Victor, and Cesar Alonso-Borrego. "Labor contracts and flexibility: evidence from a labor market reform in Spain." Economic Inquiry 52.2 (2014): 930-957.

[2] Baltagi, Badi H., and Yin-Fang Yen. "Hospital treatment rates and spillover effects: Does ownership matter?." Regional Science and Urban Economics 49 (2014): 193-202.

[3] Battisti, Michele, Giovanni Peri, and Agnese Romiti. "Dynamic Effects of Co-Ethnic Networks on Human Capital and Employment of Immigrants." NBER Working Paper #22389 (2016).

[4] Battu, Harminder, Paul Seaman, and Yves Zenou. "Job contact networks and the ethnic minorities." Labour Economics 18.1 (2011): 48-56.

[5] Bayer, Patrick, Stephen L. Ross, and Giorgio Topa. "Place of Work and Place of Residence: irregular Hiring Networks and Labor Market Outcomes." Journal of Political Economy 116.6 (2008): 1150-1196.

[6] Becker, Jeffrey, and Manfred Elfstrom. "The impact of China's Labor Contract Law on workers." International Labor Rights Forum, May 12 (2010).

[7] Bertrand, Marianne, Erzo FP Luttmer, and Sendhil Mullainathan. "Network Effects and Welfare Cultures." The Quarterly Journal of Economics 115.3 (2000): 1019-1055.

[8] Booth, Alison L., Marco Francesconi, and Jeff Frank. "Temporary jobs: stepping stones or dead ends?." The Economic Journal 112.480 (2002): F189-F213.

[9] Calvo-Armengol, Antoni, and Matthew O. Jackson. "The effects of social networks on employment and inequality." The American Economic Review 94.3 (2004): 426-454.

[10] Carrington, William J., Enrica Detragiache, and Tara Vishwanath. "Migration with endogenous moving costs." The American Economic Review 86.4 (1996): 909-930.

[11] Chen, Yuyu, Jin G. Zhe, and Yue Yang. "Peer Migration in China." NBER Working Paper #15671 (2010).

[12] Cingano, Federico, and Alfonso Rosolia. "People I know: job search and social networks." Journal of Labor Economics 30.2 (2012): 291-332.

[13] Cliff, Andrew, and Keith Ord. Spatial Autocorrelation. Pion, London, 1973.

[14] Duflo, Esther, and Emmanuel Saez. "The Role Of Information And Social Interactions In Retirement Plan Decisions: Evidence From A Randomized Experiment." The Quarterly Journal of Economics 118.3 (2003): 815-842.

[15] Elliott, James R. "Referral hiring and ethnically homogeneous jobs: How prevalent is the connection and for whom?." Social Science Research 30.3 (2001): 401-425.

[16] Falcón, Luis M. "Social networks and latino immigrants in the labor market: A review of the literature and evidence." Latinos in a Changing Society 11 (2007): 254-272.

[17] Frijters, Paul, Michael A. Shields, and Stephen Wheatley Price. "Job search methods and their success: a comparison of immigrants and natives in the UK." The Economic Journal 115.507 (2005): F359-F376.

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

[19] Kelejian, Harry H., Ingmar R. Prucha, and Yevgeny Yuzefovich. "Instrumental variable estimation of a spatial autoregressive model with autoregressive disturbances: Large and small sample results." Advances in Econometrics: Spatial and Spatio-Temporal econometrics (2004): 163-198.

[20] Hellerstein, Judith K., Melissa McInerney, and David Neumark. "Measuring the importance of labor market networks." IZA Discussion Paper 3570 (2008).

[21] Ioannides, Yannis M., and Linda Datcher Loury. "Job information networks, neighborhood effects, and inequality." Journal of Economic Literature 42.4 (2004): 1056-1093.

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

[23] Lee, Lung-fei. "Best Spatial Two-Stage Least Squares Estimators for a Spatial Autoregressive Model with Autoregressive Disturbances." Econometric Reviews 22.4 (2003): 307-335.

[24] LeSage, James P. "Bayesian estimation of limited dependent variable spatial autoregressive models." Geographical Analysis 32.1 (2000): 19-35.

[25] LeSage, James, and Robert Kelley Pace. Introduction to Spatial Econometrics. CRC Press, 2009.

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

[27] Loury, Linda Datcher. "Some contacts are more equal than others: Informal networks, job tenure, and wages." Journal of Labor Economics 24.2 (2006): 299-318.

[28] Lovett, Steve, Lee C. Simmons, and Raja Kali. "Guanxi versus the market: Ethics and efficiency." Journal of International Business Studies 30.2 (1999): 231-247.

[29] Marmaros, David, and Bruce Sacerdote. "Peer and social networks in job search." European Economic Review 46.4 (2002): 870-879.

[30] McMillen, Daniel P. "Probit with spatial autocorrelation." Journal of Regional Science 32.3 (1992): 335-348.

[31] Montogomery, James. "Social networks and labor market outcomes." American Economic Review 8.5 (1991): 1408-1418.

[32] Munshi, Kaivan. "Networks in the modern economy: Mexican migrants in the US labor market." The Quarterly Journal of Economics (2003): 549-599.

[33] Patacchini, Eleonora, and Yves Zenou. "Ethnic networks and employment outcomes." Regional Science and Urban Economics 42.6 (2012): 938-949.

[34] Pinkse, Joris, and Margaret E. Slade. "Contracting in space: An application of spatial statistics to discrete-choice models." Journal of Econometrics 85.1 (1998): 125-154.

[35] Smirnov, Oleg A. "Modeling spatial discrete choice." Regional Science and Urban Economics 40.5 (2010): 292-298.

[36] Zhang, Xiaobo, and Guo Li. "Does guanxi matter to nonfarm employment?." Journal of Comparative Economics 31.2 (2003): 315-331.