The Impact of Job Contact Networks on Wages of Rural-Urban Migrants in China: A Switching Regression Approach[*]

Wenjin Long[a], Simon Appleton [b], Lina Song [cd]

[a] **Wenjin Long**

Nottingham University Business School
Nottingham, NG8 1BB, UK
longwenjin@gmail.com, +86 13401082313

[b] **Simon Appleton**

School of Economics
The University of Nottingham, Nottingham, NG7 2RD, UK
simon.appleton@nottingham.ac.uk, +44 115 8466105

[c] **Lina Song,** [d] **Corresponding author**

Division of Industrial Economics and Finance
Nottingham University Business School, Nottingham, NG8 1BB, UK
lina.song@nottingham.ac.uk, +44 115 8466217

[*]**Acknowledgement**

ESRC Grants ES/1035079/1 and ES/M004333/1 are gratefully acknowledged.

---

[*] The paper is the revised version based on IZA DP No. 10139.

The Impact of Job Contact Networks on Wages of Rural-Urban Migrants in China: A Switching Regression Approach

ABSTRACT

In nationally representative household data from the 2008 Chinese Rural to Urban Migration Survey, nearly two thirds of rural-urban migrants found their employment through family members, relatives, friends or acquaintances. This paper investigates why the use of social network to find jobs is so prevalent among rural-urban migrants in China, and whether migrants face a wage penalty as a result of adopting this job search method. Using a switch regression approach, we find evidence of positive selection effects of the use of networks on wages. Users of networks tend to be older, to have migrated longer ago and to be less educated. In addition, married workers and those from villages with more out-migrant are more likely to use networks, while those without local residential registration status are less likely. Controlling for selectivity, we find a large negative impact of network use on wages. Using job contacts brings access to urban employment, but at the cost of markedly lower wages.

JEL Classification: J24, J31, O15

KEYWORDS: Social network; Job contact; Wage; Rural-urban migrant; Switching regression

1. Introduction

There has been increased interest in the role of social networks and its relation to the economic outcomes. Following Coleman (1988), 'social capital' has been hypothesised as a determinant of productivity, much as human capital was earlier added to the conventional factors of production like land, labour and capital. Social capital has variously been understood as norms of behaviour such as 'trust' that may underpin economic relations or as 'networks' of contacts that may provide valuable economic gains. Early empirical studies often found large economic benefits to social capital, for example, for household income in rural Tanzania (Narayan and Pritchett 1999); for manufacturing productivity in Ghana (Barr 2000); or agricultural traders in Madagascar (Fafchamps and Minten 2002). The significant role of job contacts in obtaining employment has long been recognised, although what is less well understood are the possible effects on subsequent wages or labour productivity of using such networks. In this paper, we estimate the effect of using social networks on the wages of migrants in China.

Much as the general literature on social capital focused on productivity benefits, theoretical models of labour markets often predict beneficial effects on wages from using job contacts. In line with the 'social capital as networks' interpretation, using job contacts may give job searchers more information about opportunities and ensure a better matching of workers to jobs, translating into higher wages for users of networks

(Montgomery 1991; Mortensen and Vishwanath 1994). Additionally, following the 'social capital as trust' approach, those who find jobs by using contacts may feel additional peer pressure to perform and thus attain higher productivity and wages (Kandel and Lazear 1992). However, while positive wage effects from using job contacts are sometimes found, this is far from universal and has led to consideration being given as to why using job contacts may appear to lower wages (Delattre and Sabatier 2007). One explanation centres on training costs (Pellizzari 2010). Firms may only be able to expend extra effort in using formal means of filling posts (for example, advertising or using recruitment agencies), rather than informal means (using social networks). Consequently, the use of formal means of filling posts will be more common where the costs of the posts remaining unfilled are high, as is likely to be the case with posts with high training costs (and consequent high wages). A second explanation centres on job seeker impatience: those keen to find employment quickly may use job contacts, sacrificing potentially higher wages from better matched posts for quicker entry into work (Bentolila, Michelacci, and Suarez 2010).

Given the ambiguous theoretical predictions about the impact of using job contacts on subsequent wages, it is interesting to try to estimate the impact empirically. However, while it is straightforward to measure the correlation between using of job contacts and subsequent wages, one must be cautious about drawing causal inferences from this due to unobservables that may affect both uses of social networks and labour market outcomes (Mouw 2006). Users of job contacts to find employment typically differ from non-users in various observed characteristics for example, they may be often older. If they are found to have higher (or lower) wages than non-users, the suspicion arises that this may simply reflect higher (or lower) unobserved productivity. In order to control for the possible selectivity of use of job contacts, we employ a switching regressions model for wages.

We focus on the impact of using social network on labour market outcomes for rural-urban migrants in China. This is an interesting case to study, not only because rural-urban migration in China is the largest human movement in the history of the world in terms of quantity, but also because rural-urban migrants in China cannot access institutional assistance from governments at their destinations. They are often forced to rely on using their social contacts to find work but the effect of this on labour market outcomes in China is not yet well understood.

Our intention in this study is therefore to answer three main questions. First, what are the determinants of migrants using contacts finding employment? Second, is the use of job contacts endogenous to wages: that is to say, are the unobservables which influence job search methods correlated to those which determine wages? Third, what is the causal effect of using job contacts on subsequent wages: do users of networks enjoy higher or lower wages than non-users, ceteris paribus? These questions may have implications for policy - for example, people would question the existence of public employment agencies if social networks have positive impacts on job search and wages (Delattre and Sabatier 2007). Conversely, in the Chinese context, if using networks appears to incur a wage penalty, it may spur policymakers to find ways to increase rural people's access to formal means of finding urban jobs.

The remainder of the paper is structured as follows. Section 2 reviews the literature, focusing on the measures of social networks and the methods to identify their effects on labour market outcomes. Section 3 describes the data and the switching regressions model to be estimated. Section 4 reports descriptive statistics and Section 5 presents the econometric results. Section 6 concludes.

2. Literature review

There are a great number of studies on the relationship between social networks and labour market outcomes. Studies may disagree because they analyse different contexts or use different datasets. They may also disagree for two methodological reasons. First, studies use different proxies to measure social networks. Second, studies employ different econometric models to estimate the relation between social networks and labour market outcomes. Consequently, in this section, we will discuss alternative measures of social network and the econometric methods commonly used in the existing literature.

2.1 Defining and measuring social networks

Recent literature has largely focused on econometric methods and tended to neglect the issue of the measurement of social networks. Obtaining agreement on suitable econometric methods only but ignoring the definition of social network may lead to incomparable estimates. Our intention therefore is to pay attention to the measurement of social networks to avoid contradictory estimates resulting from different measures of social network when examining the relationship between social networks and their labour market outcomes.

Guanxi (social relations in Chinese) has been largely equated to social networks by many scholars. Zhang and Li (2003) categorise three types of guanxi for migrants: receiving help from relatives or family members during the process of finding a job; having contacts outside the dwelling region or receiving remittances transactions from non-family members; and having family member who are cadres.

Some researchers use more direct measurements of social network. For example, Giulietti *et al*. (2010) measure social network by including both the quality and quantity of contacts. To them, the size of the social network is calculated by the total number of people living in cities whom they had sent greeting cards to during the last Spring Festival, traditionally the most celebrated holiday for rural Chinese. The quality of social network is proxied by a dummy variable which equals to 1 if the closest supporter is currently employed. Mouw (2010) uses the number of relatives in the U.S. to measure Hispanic workers' social capital. Wahba and Zenou (2005) use population density as a proxy for the size of social network and use the unemployment rate as a proxy for the quality of social network.

The most relevant measurement to this study is in Chen (2011), who reviews three kinds of social network: 1) the treatment effects of using contacts or social networks on individual's labour market outcomes; 2) the quantity and quality of public and social resources already mobilized from a given helper; 3) the quantity and quality of the public and social resources to which an individual has access to. This study measures social network adopting the first of Chen's three definitions: relying on personal contacts to find jobs in the labour market. This is the most common measure of social networks used in labour market studies and so places us within the mainstream of work on this issue.

2.2 Econometric methods for exploring the relationship between social networks and wages

The most common method to explore the relationship between social network and wages is to run an OLS wage equation with a variable that measures social networks. Previous studies include Aguilera (2005); Granovetter (1973, 1983); Montgomery (1991); Bian (1997) and Zhang and Li (2003). For example, we can estimate a Mincer wage equation and include a dummy variable that equal to 1 when people use their social network to find a job. The coefficient of this dummy variable captures the level effect of social network. An assumption of this method is the restriction of the coefficient to being the same between users of social network and non-users of social network, which might be wrong in reality.

Another problem with OLS is that there might be sample-selection bias on unobserved abilities in choosing job search methods. For example, migrants with high skills may be more likely to choose market-methods. A simultaneity bias may be raised if the use of social network to find a job is correlated with the expectation of labour market outcomes (Giulietti, Schlutery, and Wahba 2011). Omitted variables that are positively correlated with both social network and earnings may lead to overestimation of the coefficient of networks. Therefore, endogeneity must be addressed and several methods have been applied to do this.

Random assignment including social and natural experiments is the best way to correct self-selection bias (Mouw 2006). For example, Beaman and Magruder (2012) create short-term jobs in a laboratory to investigate who gets job referrals. They ask every participant to refer a friend who is most suitable for the job. Thus the type of referral contract and amount offered is randomised and selection bias is avoided. Khan and Lehrer (2013) try to identify the effects of changes of social networks on employment by using the random assignment method.

Using panel data is another way to eliminate unobserved fixed effects (Battu, Seaman, and Zenou 2011; Yakubovich 2005). However, the use of panel data may depend on the definition of social network. For example, Knight and Yueh (2008) use Chinese Communist Party membership and the number of close contacts of the respondent as measures of social capital. They admit that panel data might fail to explore the effect of social capital on the labour market as CCP membership and the number of close contacts does not vary much over time or may even be time-invariant for many observations.

A third approach to estimating the effect of networks on labour market outcomes is to model the selection process and build a model of labour market outcomes that controls for the selectivity. Instrumental variable methods (IV), treatment effect model, Heckman's Selection model, endogenous switching regression model, propensity score matching (Ye et al. 2012), and structural equation models (SEM) are typical methods to discover the selection process under restricted data.

With IV and related econometric methods, the practical problem is finding an appropriate instrumental variable. Distance from home village to destination might be a potential instrumental variable for social networks. However, Zhang and Zhao (2013) prove that rural-urban migrant incomes are correlated with the distance from home village to destination as migrants face an income-distance trade-off. Knight and Yueh (2008) admit that even with the use of the best instruments available in the dataset they cannot deny that social network and CCP membership may be correlated with unobserved factors. The endogeneity still needed to be addressed.

Both treatment effects model and Heckman selection models are two-stage processes. A two-step strategy involves first using a probit model on the complete data. They then calculate an inverse Mills ratio (non-selection hazard rate) for selected samples and selection hazard rate for unselected samples. The hazard rate variable is then used as an additional regressor in the second stage OLS model. They can be estimated by either the maximum likelihood method or a two stage least squares method. The difference between Heckman selection model and treatment effects model is that the former only uses the samples which are observed and does not include a dummy variable in wage equation, whereas, the latter uses the full sample and includes a dummy variable in wage equation Guo and Fraser (2009). The Heckman selection model is used in Amuedo-Dorantes and Mundra (2007) and Yueh (2008).

The switching regression approach uses the Heckman two-stage procedure twice, one for selected sample and another for unselected sample. Studies using switching regressions can be found in Delattre and Sabatier (2007) and Liang (2010). Dutoit (2007) suggests that Hackman's selection model should be more appropriate when one regime is missing, whereas, switching regression model is more appropriate when both regimes are observed. Specifically, the endogenous switching regression is used to address issues of self-selection and the estimation of treatment effects when the self-selection is not random and the control group is observed.

In this paper, we use the switching regression approach as the data provides information of both users and non-users of social networks, and we need to address issues of self-selection bias in investigating the effect of social networks on wages.

3. Data and methodology

3.1 Data

The empirical analysis is based on the first wave of the Rural to Urban Migration in China survey (RUMIC 2008). This survey was conducted in 2008 and includes approximately 5,000 rural-urban migrant households. Samples were randomly chosen in the fifteen migrant destination cities nationally. The dataset includes detailed information about social-demographic characteristics, labour market outcomes, health conditions, major life events and social network information. Our analysis is based on 5453 employed migrant workers, aged 16-60 years old.

Our focus in this paper is on real wages, defined to include bonuses, allowances and income in-kind. Results from the survey are reported in Dmurger and Li (2012), Qu and Zhao (2011) and Meng, Kong, and Zhang (2010).

Granovetter (1974) notes that individuals may use several methods to find a job and only one method leads to success. But the dataset we use only ask the main source of job finding, i.e., the one leads to retaining the job. The dataset does not have variables about information flows between employers, contacts, and employees, so we are limited in what we can analyse about the process of job search.

3.2 Methods: Endogenous Switching Regression Model

An endogenous switching regression is used to estimate the relation between using of social network and wages, controlling for self-selection bias. The endogenous switching regression can be used to predict expected wages for users of social networks if they switched to not using networks and vice versa for non-users (Dutoit 2007; Lokshin and Sajaia 2004; Powers 2007). This method has proved to be useful in dealing with endogeneity from self-selection by Adamchik and Bedi (2000) and Heitmueller (2006). Different decomposition methods can be used after calculating consistent estimates. OLS regression and treatment effects models will also be estimated to give comparisons and gauge robustness.

Let Z^* denote a latent variable for the propensity to use social networks with the following index function:

$$Z^* = W\gamma + \upsilon, \upsilon \sim N(0, \sigma_\upsilon^2)$$

$$Z = 1 \; if \; Z^* > 0, \qquad Z = 0 \; otherwise$$

where Z=1 if the individual chooses social networks to find a job and Z=0 otherwise. Z^* can be estimated using models for binary data.

Let Y_1 be the earnings for users of social networks and let Y_0 be the earnings for non-users of social networks.

$$Y_1 = X_1\beta_1 + \mu_{1,} \; \mu_1 \sim N(0, \sigma_1^2) \; if \; Z = 1$$

and

$$Y_0 = X_0\beta_0 + \mu_{0,} \; \mu_0 \sim N(0, \sigma_0^2) \; if \; Z = 0$$

In practice, we observe sample respondents in only one state Z = 1 or Z = 0. That is, we observe Y_1 when Z = 1, in which case Y_0 is unobserved. Similarly, we observe Y_0 when Z = 0, in which case Y_1 is unobserved.

Assume that υ, μ_1, μ_0 have a trivariate normal distribution with mean vector zero and covariance matrix:

$$\Omega = \begin{bmatrix} \sigma_\upsilon^2 & \sigma_{1\upsilon} & \sigma_{0\upsilon} \\ \sigma_{1\upsilon} & \sigma_1^2 & \cdot \\ \sigma_{0\upsilon} & \cdot & \sigma_0^2 \end{bmatrix}$$

where σ_{υ}^{2} is a variance of the error term in the selection equation, normalising with $\sigma_{\upsilon}^{2}=1$. σ_{1}^{2} and σ_{0}^{2} are variances of the error terms in the wage equations. $\sigma_{1\upsilon}$ is a covariance of μ_{1} and υ, and $\sigma_{0\upsilon}$ is a covariance of μ_{0} and υ. The covariance between μ_{1} and μ_{0} is not defined, as Y_1 and Y_0 are never observed simultaneously.

Using direct maximum likelihood methods, the log-likelihood function associated with our model is written as

$$\text{Log L}=\sum_{Z=1}\left[\log\Phi(\eta_{1})-\frac{1}{2}\left\{\log(2\pi\sigma_{1\upsilon})-\left(\frac{Y-X\beta_{1}}{\sigma_{1}}\right)^{2}\right\}\right]$$
$$+\sum_{Z=0}\left[\log\Phi(\eta_{0})-\frac{1}{2}\left\{\log(2\pi\sigma_{0\upsilon})-\left(\frac{Y-X\beta_{0}}{\sigma_{0}}\right)^{2}\right\}\right]$$

where

$$\eta_{1}=\frac{W\gamma+(Y-X\beta_{1})\sigma_{1\upsilon}/\sigma_{1}^{2}}{\sqrt{1-\rho_{1\upsilon}^{2}}}\quad\text{and}\quad\eta_{0}=\frac{W\gamma+(Y-X\beta_{0})\sigma_{0\upsilon}/\sigma_{0}^{2}}{\sqrt{1-\rho_{0\upsilon}^{2}}}$$

Note: $\rho_{0\upsilon}=\frac{\sigma_{0\upsilon}^{2}}{\sigma_{0}\sigma_{\upsilon}}=\frac{\sigma_{0\upsilon}}{\sigma_{0}}$ is the correlation coefficient between μ_{0} and υ.

$\rho_{1\upsilon}=\frac{\sigma_{1\upsilon}^{2}}{\sigma_{1}\sigma_{\upsilon}}=\frac{\sigma_{1\upsilon}}{\sigma_{1}}$ is the correlation coefficient between μ_{1} and υ.

$\phi(\cdot)$ is the standard normal density function.

$\Phi(\cdot)$ is the standard normal cumulative distribution function.

After estimating the models parameters, the following unconditional expectations could be calculated:

$$\mathrm{E}(Y_{1}|\mathrm{Z}=1)=X_{1}\beta_{1}\qquad(1)$$

$$\mathrm{E}(Y_0|\mathrm{Z}=0) = X_0\beta_0 \qquad (2)$$

Because of the selection problem (the failure to observe Y_0 when Z = 1 and the failure to observe Y_1 when Z = 0), we need to write these outcomes in a selection-equation format. Taking expectations of the outcome equations, we can find the conditional expected earnings for a user of social networks who self-selected into user of social networks as follows:

$$\mathrm{E}(Y_1|\mathrm{Z}=1,\ X_1) = E(Y_1|Z^* > 0)$$

$$= E(Y_1|W\gamma + \upsilon > 0)$$

$$= X_1\beta_1 + E(\mu_1|\upsilon < W\gamma)$$

$$= X_1\beta_1 + \sigma_1\rho_{1\upsilon}\frac{\phi(W\gamma)}{\Phi(W\gamma)} \qquad (3)$$

Where $\frac{\phi(W\gamma)}{\Phi(W\gamma)}$ is referred to as non-selection hazard rate or the inverse Mills ratio.

Similarly, the conditional expected earnings for a non-user of social networks who self-selected into non-user of social networks is:

$$\mathrm{E}(Y_0|\mathrm{Z}=0,\ X_0) = E(Y_0|Z^* < 0)$$

$$= E(Y_0|W\gamma + \upsilon < 0)$$

$$= X_0\beta_0 + E(\mu_0|\upsilon > W\gamma)$$

$$= X_0\beta_0 - \sigma_0\rho_{0\upsilon}\frac{\phi(W\gamma)}{1-\Phi(W\gamma)} \qquad (4)$$

Where $\frac{\phi(W\gamma)}{1-\Phi(W\gamma)}$ is referred to as the selection hazard rate.

In experimental studies, those assigned to status 1 differ only randomly from those assigned to status 0, so that there is an interchangeability across statuses. However, in non-experimental studies one must adjust for selectivity when evaluating expected earnings if individuals were assigned to statuses other than the ones they in fact entered. The expected outcome for users of social networks if they were non-users is:

$$\mathrm{E}(Y_1|\mathrm{Z}=0,\ X_1) = E(Y_1|Z^* < 0)$$

$$= E(Y_1|W\gamma + \upsilon < 0)$$

$$= X_1\beta_1 + E(\mu_1|\upsilon > W\gamma)$$

$$= X_1\beta_1 - \sigma_1\rho_{1\upsilon}\frac{\phi(W\gamma)}{1-\Phi(W\gamma)} \qquad (5)$$

Similarly, the expected outcome for non-users of social networks, had they been users, is:

$$\mathrm{E}(Y_0|\mathrm{Z}=1,\ X_0) = E(Y_0|Z^* > 0)$$

$$= E(Y_0|W\gamma + \upsilon > 0)$$

$$= X_0\beta_0 + E(\mu_0|\upsilon < W\gamma)$$

$$= X_0\beta_0 + \sigma_0\rho_{0\upsilon}\frac{\phi(W\gamma)}{\Phi(W\gamma)} \qquad (6)$$

We can use Stata command ***movestay*** to estimate this endogenous switching regression with consistent standard errors by using the full-information maximum likelihood method (Lokshin and Sajaia 2004).

4. Descriptions

4.1 Job Search Methods

Employed individuals in the survey were asked about the way through which they obtained the current (main) jobs. The survey provides an extensive list of job search methods as listed in Table 1.

The job search methods in Table 1 can be grouped into four types. Group One is through the government, including: 1) assigned by the government, 2) through an employment agent run by government, 3) through a community employment service station; Group Two is through the market, including: 1) through a commercial employment agent run by private (including job fair), 2) applied for advertised job, 3) applied directly, 4) employer recruitment. Group Three is through social networks, including: 1) introduced by family members, 2) introduced by relatives, 3) introduced by friends, 4) introduced by acquaintances. Group four is through any other method not categorised [1].

[Table 1]

Details from Table 1 reveal that the dominant method of obtaining jobs is via migrants' social network. 31.1 percent of the respondents obtained the current jobs through friends, the most frequent means of getting jobs. The second most common job search method is through relatives, which accounts for 21 percent. 7.57 percent of them used family members and 2.92 percent used acquaintances.

When aggregating all sources of job search into the four main categories (in

Table2), we find 63% of migrants obtained their current jobs through social networks, followed by 32% through market mechanisms and only 2% through government means.

[Table 2]

To simplify the analysis but highlight on social networks, we merge those who got jobs by using government, market and others methods into one group to contrast the other group using social networks for job obtaining. This is to say, in further analysis, we only deal with two categories - the users of networks and others. Table 3 shows the details of grouping.

[Table 3]

4.2 Homophily

The descriptive data in Tables 1, 2 and 3 reveal an observed tendency of homophily - ‘like to associate with like’ in job seeking for rural-urban migrants. Contrasted with 63% of migrants getting jobs through social networks and 33% from market mechanism, only 2% utilised the government job recruitment services. The difference among the recruitment methods cannot be further explored by this research, as we do not have information on whether migrants who use social networks to get jobs also assist fellow migrants to find jobs. To be more specific, we have no information on the composition of migrants’ job contact network apart from the relationship between the job-seekers and the helpers; therefore we cannot pair the users and those who offer the help. With this, we rely on the fact that a migrant’s family members, relatives or friends are more likely to share similar status.

Homophily does matter in this research if we regard the government recruitment service as a public good yet find that it hardly serves rural-urban migrants, the big group of Chinese labour force. We further learn from Table 4 that it is the old and less educated are more likely to use social networks for job obtaining, and less so for the young and educated (Table 4). Does this reveal a restriction for them to get efficient information on jobs or simply a lack of market demand of them?

Due to data limitations, we cannot detect whether the large migrant cohort has taken actions of such is based on economic rationality or just a series of individual events due to the lack of information or resources. However, we can test how their earnings are affected by different job-search methods. Thus, we employ a rigorous econometric approach to establish the relationship between the earnings and job-obtaining methods.

[Table 4]

Table 4 highlights the extent to which age and education are correlated with the use of social networks. Older and less educated people are more likely to use social networks to find jobs while young and educated people tend to find jobs through the market, the government and other channels. This shows that the use of social networks is not randomly distributed across individual characteristics and thus may be endogenous to wages. Gender differences between the users and non-users of network are small. Married people are more likely to use social network than unmarried people. All these variables are included in our further tests.

4.3 Job Search Methods and Wages

Migrants who use social networks receive lower wages than those who do not. Table 5 reports mean hourly wages for users and non-users of social networks. The gap in hourly wages is 0.64 yuan or 11.7%. This suggests that there is wage cost to using social networks, although further analysis is required to establish a causal impact.

[Table 5]

We plot the logged hourly wage using the kernel density graph (Figure 1) to show how the distribution of earnings differs between the two types of job-seekers. We see that plot of the log hourly wage for the users of networks is slightly to the left of that for non-users, showing a gap across the distribution.

[Figure 1]

5. Models for Job Search Methods and Wages

5.1 OLS regressions

Our estimations start with OLS regressions on wage rate. The explanatory variables we control for are divided into personal characteristics, job related characteristics and a set of dummy variables for provinces. Personal characteristics include gender, age, age squared, years of migration, years of migration squared, and years of schooling. Job related characteristics include occupation, industry, ownership sector and firm size. Our focus is on whether using networks to find jobs has an impact on the wages subsequently earned on that job. From the OLS regressions in Table 6, the magnitude of the coefficient is small, implying the use of social networks is associated with 1% lower wages, ceteris paribus. Moreover, the coefficient is not statistically significant; using OLS, we cannot reject the null hypothesis that network use has no effect on wages.

[Table 6]

5.2 Switching regression

Since OLS estimates may be biased due to self-selection over the use of social networks, we run a switching regression instead. The key issue in controlling for the endogeneity of social networks is identification: finding instrumental variables that could be included in the selection model and excluded from the wage equation. Delattre and Sabatier (2007) use parental occupation and ease of access to public employment agencies to identify the effect of using social networks on wages in France [2].

From the variables we have in the dataset, we use marriage status, ratio of migrants in the home village and hukou status as identifying variables. Married workers may have access to a wider network of contacts, via their spouses, encouraging the use of networks. Ye and Zhou (2010) suggest that marriage does not affect migrant earnings directly because most migrants are working on less skilled jobs. Where there is a large proportion of rural-urban migrants being sent from the worker’s home village, workers in that village are likely to have more contacts who can help them find urban jobs. Hence we follow Chen (2009) and Chen, Jin, and Yue (2010) in using the proportion of labour migrants in the home village as an indicator of the village social network. When the household registration (hukou) of the worker is not local, they may have fewer contacts with local people and thus less chance of using their networks to find employment. We assume that these identifying variables influence the selection choice to use social networks but not impact wages directly (assumptions that we test empirically).

The selection model from Table 7 shows that each of three hypothesised identifying variables have significant effects on the likelihood of using social networks to find employment, with the predicted signs. Amongst the other control variables, the use of social networks declines with education level and increases with age and with years of migration. Older workers are likely to have accumulated more acquaintances and thus have access to a wider social network. Educated workers may have less need to use social networks to find jobs, as their qualifications give them an edge in job competition. Age, age squared, years of migration, years of migration squared, and educational level are included in selection model.

[Table 7]

Table 8 shows the results of wage equations from the switching regressions. The rho values in Table 8, which measure the correlation coefficients between the error terms in the selection model and the wage equation, are both statistically significant.

This implies that the use of social network is endogenous to wage determination. Thus the switching regression is more appropriate than the OLS regression. The rho term of non-users of networks, corresponds to $\rho_{0\upsilon}$ discussed earlier: its positive value implies that unobserved factors υ which increase the likelihood of using social networks are positive correlated with the unobserved determinants μ_0 of the wages of those not using networks. Similarly, the positive value of the rho term for users of networks, corresponds to $\rho_{1\upsilon}$ and implies υ is positively correlated with the unobserved determinants μ_1 of the wages of those using networks.

[Table 8]

There is thus positive selection in network use - those who use social networks are likely to be paid more due to unobservable factors. Note that this positive selectivity is present in both regimes - users of networks have unobservables that would imply higher wages conditional on using networks, but also conditional on not using networks. This implies that users of networks tend to have unobserved characteristics ('ability') that raise wages, or perhaps to work in jobs with unobserved characteristics that are associated with higher pay. Not controlling for this, for example in the OLS models, will bias upwards estimates of the impact of network usage on wages. Delattre and Sabatier (2007) obtained the same finding when analysing data from France in 1995.

The switching regression model differs from an endogenous dummy variable model in allowing explanatory variables to impact wages differently between users of network and non-users of network. However, comparing the wage equations for users of networks and for non-users, few differences are apparent. Generally speaking, the determinants of wages appear similar in the two sub-samples. Wald tests for differences in the individual coefficients reveal no differences in the explanatory variables that are statistically significant at the 5% level. The coefficients on age and age squared are each significantly different in the two samples at the 10% level. Plotting the joint effects of the age quadratics, it is apparent that wages fall with age at a faster rate for users of networks than for non-users. The constant terms in the two wage equations are significantly different from each other at the 1% level, implying wage differences between the two sub-samples.

5.3 Wage difference

Table 9 compares the predicted log wages from the wage equations in Table 7 with the observed values. Unconditional log wages are predicted from the switching model using Equation 1 and 2. Conditional log wages are predicted from the switching model using Equation 3, 4, 5 and 6. The predicted unconditional log wages are the most useful for calculating the effect of using social networks on wages, since they are not subject to any selectivity bias.

[Table 9]

Non-users of networks have somewhat higher observed mean log wages than users, implying a wage differential of around 9%. However, the gap in unconditional predicted mean wages is much greater, consistent with a wage differential of around 59%. This reflects the positive selectivity of network use - network users have unobservable factors that raise their wages and thus the observed wage gaps masks the extent of the negative effect of using networks to find jobs on wages.

It is not uncommon in the literature to find wage discounts from using social networks. However, what is striking about our results is the magnitude of the effect. For example, Delattre and Sabatier (2007) find a 7% wage discount in France in 1995. Bentolila, Michelacci, and Suarez (2010) report discounts of at least 2.5% in the EU and US. While of the same sign, the wage differentials we estimate for rural-urban migrants in China are of an order of magnitude higher than those found in industrialised countries. We can only speculate on why the wage discount from using social networks is so much higher in China, but one factor is likely to be the large rural-urban gap in labour productivity in China. Potential rural-urban migrants may be faced with a choice between staying in low return agriculture and finding a much more remunerative non-agricultural job if they migrate. It is not uncommon to find estimates of labour productivity outside of agriculture being ten times as high as in agriculture (Knight and Song 2003). In such a situation, it is quite plausible that rural workers will accept a sizable urban wage discount from using networks, if that gives them access to urban jobs.

5.4 Robustness tests

OID test

In our selection model, we use three instrument variables: marriage status, ratio of labour migrants in home village and hukou status. We assume that the instrumental variables only affect wages through network use. We test this assumption by an over-identifying test for instrumental variables in Table 10. The dependent variable is the residual of wage equation from the switching regressions. The independent variables are all the exogenous regressors in the wage equations together with the three instrumental variables. This auxiliary regression shows that all the instrumental variables do not affect the residuals of the wage equations, so we cannot reject our assumption that they can be excluded from the wage equations.

[Table 10]

In preliminary work, we explored various other potential instruments for network usage, including family political background, parental education, parental occupation, number of siblings, number of friends living in urban areas and the distance from home province to working province. However, these were rejected as instruments either because they were not correlated with network user or because they were correlated with wage residuals from the switching regressions. Specifically, although family political background is a good IV in Zhang and Lu (2009), it is not correlated with the use of networks in our data. Likewise, parents' education and occupation are not significantly correlated with the use of network. The number of siblings, the number of friends living in city and the distance from the home province to the workplace province are correlated with the use of network, but fail the OID test.

Treatment effects models

An alternative to the switching regressions approach is to use a treatment effects model, in which use of social networks in job search is included as an endogenous dummy variable in a wage equation. Both Maximum Likelihood (ML) and Two Step estimators imply the dummy variable has large, significant negative impacts on wages. The coefficient implies use of networks would lower log wages by 0.438 points. This is of a similar order of magnitude to that implied by a comparison of the predicted unconditional mean log wages from the switching regression models.

[Table 11]

6. Conclusion

It is well understood that social networks play an important role in labour markets, with many people finding jobs through their friends, acquaintances and relatives. In our survey of rural-urban migrants in China in 2008, nearly two thirds found their employment by such job contacts. However, what is less clear is the impact of such job search mechanisms on subsequent wages. Addressing this issue in a convincing manner requires explicit consideration of the selection process which determines whether or not workers use social networks to find their jobs and whether this causes any selectivity bias for wages.

In this paper, we modelled the selection of job search methods. We found that users of networks to find jobs tended to be older, to have been migrants for longer and to be less educated. In addition, married workers and those from villages which sent out many migrants were more likely to use networks, while those with non-local household registration were less likely. The last three effects - marital status, migration rates in home village and household registration – were used as instrumental variables to identify the impact of network use on wages.

We found evidence of positive selection effects between use of networks and wages. Those using networks to find their jobs appear to have unobservables that are associated with higher pay - this could be unobserved personal characteristics such as ability or perhaps remunerative job characteristics. Thus, although users of networks are less educated than non-users, they appear to have other, unobserved advantages for pay determination. This implies that simple estimations of the impact of network use on migration that do not control for selectivity are upwards biased - benefits would be ascribed to networks that more properly reflect the favourable unobserved characteristics of network users.

Controlling for selectivity, we find a large negative impact of network use on wages. Using job contacts brings benefits in terms of giving people from rural China access to urban employment. However, it also imposes a substantial cost - those urban jobs found using job contacts pay markedly less than those found by impersonal means. This negative effect of using employment has been found before, using comparable methods, in industrialised countries. However, what is striking about our results is the magnitude of the effect. This may be explicable given the greater segmentation of labour in China and the consequent large gap in labour productivity between agriculture and the employment of rural-urban migrants. At first look it appears a puzzle, why so many rural-urban migrants use friends and relatives to find jobs if they face a wage penalty as a result. However, such migrants may have less chance of finding jobs if they rely on formal means of job search. Given the large gap in labour productivity rural and urban areas, it may well be worth accepting a wage discount by using social networks if it is the means by which the migrants can escape low return farming and find more remunerative work in urban areas.

Before drawing strong policy conclusions, further research is required into the possible barriers facing migrants in accessing market or state mechanisms for urban job search. However, our paper does provide prima facie evidence of imperfections in the labour market: many migrant workers appear unable to use such formal mechanisms, so the majority rely on social networks – apparently at the cost of obtaining substantially lower paying jobs. This finding does not reflect a selectivity bias – indeed, correcting for such bias dramatically increases the estimated wage cost of using networks to find employment. Our research suggest that efforts to publicise information about urban job vacancies in rural areas, whether through expanding existing state job exchanges or promoting private alternatives, could bring large benefits for migrant workers. Not only would such efforts increase access to urban employment for migrants, they may also raise the wages paid if they do find urban jobs. Social capital may be substituting for the lack of well-functioning formal labour market institutions, but it is far from a perfect substitute.

Notes

1. Granovetter (1974) divides job search methods into three groups: formal means, personal contracts, and direction application. He argues direction application that does not use a formal or personal intermediary is different from both formal means and from personal contacts.

2. Liang (2010) use a switching regression model to estimate the effect on wages of using social networks to find jobs in eight cities in China. Their exclusion restriction is to use the square of the age at which workers obtained their jobs in the selection model but not in the wage equations. This exclusion restriction seems a questionable a priori given that they include the linear term for age at which workers obtain their jobs in both the selection and wage equations.

References

Adamchik, V. A. and A. S. Bedi (2000). Wage Differentials between the Public and the Private Sectors: Evidence from an Economy in Transition. *Labour Economics* 7(2), 203-224.

Aguilera, M. B. (2005). The Impact of Social Capital on the Earnings of Puerto Rican Migrants. *The Sociological Quarterly* 46(4), 569-592.

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

Barr, A. (2000). Social Capital and Technical Information Flows in the Ghanaian Manufacturing Sector. *Oxford Economic Papers* 52(3), 539-559.

Battu, H., P. Seaman, and Y. Zenou (2011). Job Contact Networks and the Ethnic Minorities. *Labour Economics* 18(1), 48-56.

Beaman, L. and J. Magruder (2012). Who Gets the Job Referral? Evidence from a Social Networks Experiment. *American Economic Review* 102(7), 3574-93.

Bentolila, S., C. Michelacci, and J. Suarez (2010). Social Contacts and Occupational Choice. *Economica* 77(305), 20-45.

Bian, Y. (1997). Bringing Strong Ties Back in: Indirect Ties, Network Bridges, and Job Searches in China. *American Sociological Review* 62(3), 366-385.

Chen, Y. (2009). Network Effects among Rural-to-Urban Migrants in China. *Sociology Working Paper Number 2009-06, Department of Sociology, Oxford University.*

Chen, Y. (2011). The Causal Effect of Social Capital in the Labor Market Identification Challenges and Strategies. *Chinese Sociological Review* 44(2), 25.

Chen, Y., G. Z. Jin, and Y. Yue (2010). Peer Migration in China. *NBER Working Paper 15671*.

Coleman, J. S. (1988). Social Capital in the Creation of Human Capital. *American Journal of Sociology* 94, S95-S120.

Delattre, E. and M. Sabatier (2007). Social Capital and Wages: An Econometric

Evaluation of Social Networking's Effects. *Labour* 21(2), 209-236.

Dutoit, L. C. (2007). Heckman Selection Model, Endogenous and Exogenous Switching Models, A survey. *http://works.bepress.com/laure dutoit/3/.*

Dmurger, S. and S. Li (2012). Migration, Remittances and Rural Employment Patterns: Evidence from China. *Working Papers halshs-00744438, HAL.*

Fafchamps, M. and B. Minten (2002). Returns to Social Network Capital among Traders. *Oxford Economic Papers* 54(2), 173-206.

Giulietti, C., M. Guzi, Z. Zhao, and K. F. Zimmermann (2010). Social Networks and the Labour Market Outcomes of Rural to Urban Migrants in China. *www.iza.org/conference files/ReLabEco2010/guzi_m5347.pdf* .

Giulietti, C., C. Schlutery, and J. Wahba (2011). Migrants and Social Networks: Old Ideas, Lasting Myths and New Findings. *www.iza.org/conference files/Transatlantic2011/giulietti_c5931.pdf* .

Granovetter, M. (1973). The Strength of Weak Ties. *The American Journal of Sociology* 78(6), 1360-1380.

Granovetter, M. (1974). *Getting A Job: A Study of Contacts and Careers.* Cambridge: Harvard University Press.

Granovetter, M. (1983). The Strength of Weak Ties: A Network Theory Revisited. *Sociological Theory* 1, 201-233.

Guo, S. and M. W. Fraser (2009). *Propensity Score Analysis: Statistical Methods and Applications*. Sage Publications.

Heitmueller, A. (2006). Public-private Sector Pay Differentials in a Devolved Scotland. *Journal of Applied Economics* 9(2), 295-323.

Kandel, E. and E. P. Lazear (1992). Peer Pressure and Partnerships. *Journal of Political Economy* 100(4), pp. 801-817.

Khan, A. Q. and S. F. Lehrer (2013). The Impact of Social Networks on Labour Market Outcomes: New Evidence from Cape Breton. *NBER Working Paper 18786.*

Knight, J. and L. Song (2003). Chinese Peasant Choices: Migration, Rural Industry or Farming. *Oxford Development Studies* 31(2), 123-148.

Knight, J. and L. Yueh (2008). The Role of Social Capital in the Labour Market in China. Economics of Transition 16(3), 389-414.

Liang, Y. (2010). Is Social Capital and Social Networks Useless? *Sociological Studies* (5), 50-82.

Lokshin, M. and Z. Sajaia (2004). Maximum Likelihood Estimation of Endogenous Switching Regression Models. *The Stata Journal* 4(3), 10.

Meng, X., S. T. Kong, and D. Zhang (2010). How Much Do We Know about the Impact of the Economic Downturn on the Employment of Migrants? *ADBI Working Papers 194.*

Montgomery, J. D. (1991). Social Networks and Labor-Market Outcomes: Toward an Economic Analysis. *The American Economic Review* 81(5), 1408-1418.

Mortensen, D. T. and T. Vishwanath (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(1), 79-102.

Mouw, T. (2010).The Use of Social Networks among Hispanic Workers: An Indirect Test of the Effect of Social Capital. *http://www.unc.edu/ tedmouw/papers/hispcontacts.pdf* .

Narayan, D. and L. Pritchett (1999). Cents and Sociability: Household Income and Social Capital in Rural Tanzania. *Economic Development and Cultural Change* 47(4), pp. 871-897.

Neuman, S. and R. L. Oaxaca (2004a). Wage Decompositions with Selectivity-Corrected Wage Equations: A Methodological Note. *Journal of Economic Inequality* 2(1), 3-10.

~~Neuman, S. and R. L. Oaxaca (2004b). Wage Differentials in the 1990s in Israel: Endowments, Discrimination, and Selectivity. *IZA DP No. 1362.*~~

Pellizzari, M. (2010, April). Do Friends and Relatives Really Help in Getting a Good Job? *Industrial and Labor Relations Review* 63(3), 494-510.

Powers, D. (2007). Censored Regression, Sample Selection, Endogenous Switching, and Treatment-Effect Regression Models. *http://www.oir.pku.edu.cn/umich/teaching/download/20092/CDAcoursematerials/handouts/SelectRegQ.pdf* .

Qu, Z. F. and Z. Zhao (2011). Evolution of the Chinese Rural-Urban Migrant Labor Market from 2002 to 2007. *IZA DP NO. 5421.*

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

Yakubovich, V. (2005). Weak Ties, Information, and Influence: How Workers Find Jobs in a Local Russian Labor Market. *American Sociological Review* 70(3), 408-421.

Ye, J., S. Bo, C. Liu, and Y. Zhou (2012). The Levels of Social Network and Wage of Rural-urban Migrants: From a Perspective of Indentity Model. *Economic Review* (4), 31-42.

Ye, J. and Y. Zhou (2010). Social Capital Transformation and Migrants' Wage--evidence from an Survey in Beijing. *Management World* (10), 34-46.

Yueh, L. Y. (2008). Do Social Networks Increase Labour Supply Elasticity? *Applied Economics Letters* 15(1), 5 - 10.

Zhang, J. and Z. Zhao (2013). Measuring the Income-Distance Trade-off for Rural-Urban Migrants in China. *IZA DP NO. 7160.*

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

Zhang, Y. and M. Lu (2009). Whether Rural-urban Migrants' Social Network Improve Their Salary? *Management World* (3), 45-54.

Table 1: Composition of job search methods

Job search methods	Frequency	Percent
Assigned by the government	30	0.55
Through employment agent run by government	52	0.95
Through community employment service station	37	0.68
Through commercial Employment agent run by private	316	5.79
Applied for advertised job	395	7.24
Applied directly	613	11.24
Introduced by family members	413	7.57
Introduced by relatives	1145	21
Introduced by friends	1696	31.1
Introduced by acquaintance	159	2.92
Employer recruitment	447	8.2
Other (please specify)	150	2.75
Total	5453	100

Table 2: Regrouped composition of job search methods

Job search methods	Frequency	Percent
Through the government	119	2.18
Through the market	1771	32.48
Through social network	3413	62.59
Other	150	2.75
Total	5453	100

Table 3: Users of network and non-users of network

Job search methods	Frequency	Percent
Users of network	3413	62.59
Non-users of network	2040	37.41
Total	5453	100

Table 4: Users of network and non-users of network

	Non-users of network	Users of network	Total
Female	42.25	42.43	42.36
Male	57.75	57.57	57.64
Married	52.99	60.01	57.38
Unmarried	47.01	39.99	42.62
Age 16-25	45.54	40.2	42.20
Age 26-35	28.53	28.77	28.68
Age 36-45	20.05	22.21	21.40
Age 46 and above	5.88	8.82	7.72
Elementary school or blew	10.10	14.70	12.98
Junior middle school	48.77	58.16	54.64
Senior middle school	20.83	16.87	18.36
Specialized secondary or above	20.29	10.27	14.02

Table 5: Users of network and non-users of network

	N	Hourly wage		log hourly wage	
		Mean	SD	Mean	SD
Users of network	2030	6.02	3.96	1.73	0.54
Non-users of network	3389	5.39	3.39	1.63	0.50
Total	5419	5.63	3.62	1.67	0.52

Table 6: Wage equations: OLS regressions

	All Sample		Non-users of network		Users of network	
	Coefficient	SE	Coefficient	SE	Coefficient	SE
Users of network	-0.0091	0.0135				
Age	0.0417***	0.0047	0.0294***	0.0085	0.0477***	0.0057
Age squared	-0.0647***	0.0068	-0.0481***	0.0124	-0.0729***	0.0081
Years of migration	0.0218***	0.0035	0.0316***	0.0055	0.0174***	0.0042
Years of migration squared	-0.0478***	0.0126	-0.0823***	0.0181	-0.0322**	0.0153
Male	0.1096***	0.0130	0.0841***	0.0223	0.1283***	0.0160
Years of schooling	0.0424***	0.0027	0.0500***	0.0048	0.0372***	0.0034
Occupation: Base group=White collar						
Blue collar	-0.1953***	0.0247	-0.2360***	0.0360	-0.1470***	0.0342
Other occupation	-0.1643	0.1012	-0.0748	0.1460	-0.1961	0.1330
Industry: Base group=Manufactory						
Construction	0.1204***	0.0240	0.1506***	0.0465	0.1141***	0.0281
Electricity, gas, water, IT and transportation	0.0217	0.0331	-0.0073	0.0544	0.0476	0.0418
Commerce and trade	0.0091	0.0228	0.0191	0.0413	0.0112	0.0271
Restaurant and catering	-0.1232***	0.0210	-0.1361***	0.0340	-0.1046***	0.0266
Finance, estate, health, education	-0.0374	0.0236	-0.0158	0.0395	-0.0421	0.0293
Services	-0.0626**	0.0266	-0.0636	0.0445	-0.0490	0.0331
Ownership: Base group=State and collective						
Private enterprises	-0.0477**	0.0223	-0.0724*	0.0410	-0.0367	0.0263
Self-employed Individuals	-0.0790***	0.0256	-0.0807*	0.0468	-0.0848***	0.0303
Foreign, joint venture	0.0368	0.0277	0.0240	0.0477	0.0328	0.0345
Shared company	0.0340	0.0283	0.0565	0.0495	0.0006	0.0342
Other enterprise	-0.2063	0.1403	-0.4682***	0.1528	0.1451	0.1379
Firm size: Base group= Below 8						
8-50	0.0921***	0.0192	0.0870**	0.0341	0.0925***	0.0234
Above 50	0.1347***	0.0200	0.1619***	0.0354	0.1216***	0.0241
Constant	0.7124***	0.0870	0.7912***	0.1478	0.6458***	0.1065
Adjust R^2	0.3330		0.3384		0.3283	
N	4845		1819		3026	

Notes: Dependent variable=log of hourly wage; SE refers to robust standard error; Provinces controlled. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 7: Switching regression: selection model

	Coefficient	SE
Unmarried	-0.1572***	0.0524
Non-local rural hukou	-0.1193***	0.0436
Ratio of labour migrants in home village	0.0044***	0.0008
Age	-0.0382**	0.0158
Age squared	0.0528**	0.0218
Years of migration	-0.0139	0.0113
Years of migration squared	0.0848**	0.0426
Male	0.0288	0.0395
Education: base group=Junior middle school		
Elementary school or blow	0.1265**	0.0593
Senior middle school	-0.1854***	0.0443
Specialized secondary school	-0.4888***	0.0519
Constant	0.9628***	0.2633
N	4723	

Note: Dependent variable: dummy=1 if user of network.

* p < 0.10, ** p < 0.05, *** p < 0.01

Table 8: Switching regression: wage equation

	Non-users of network		Users of network	
	Coefficient	SE	Coefficient	SE
Age	0.0255***	0.0091	0.0463***	0.0059
Age squared	-0.0412***	0.0133	-0.0700***	0.0084
Years of migration	0.0284***	0.007	0.0150***	0.0045
Years of migration squared	-0.0615**	0.0269	-0.0174	0.0161
Male	0.0923***	0.0244	0.1278***	0.0176
Years of schooling	0.0346***	0.0052	0.0288***	0.0036
Occupation: Base group=White collar				
Blue collar	-0.2137***	0.0345	-0.1376***	0.0315
Other occupation	-0.0708	0.1655	-0.1754	0.1231
Industry: Base group=Manufactory				
Construction	0.1572***	0.0425	0.1066***	0.0281
Electricity, gas, water, IT and transportation	-0.0054	0.0572	0.0506	0.0414
Commerce and trade	0.0254	0.0383	0.0076	0.028
Restaurant and catering	-0.1379***	0.0348	-0.1109***	0.0268
Finance, estate, health, education	-0.025	0.0378	-0.0437	0.0295
Services	-0.0599	0.0393	-0.0588*	0.0310
Ownership: Base group=State and collective				
Private enterprises	-0.0501	0.0399	-0.0411	0.0261
Self-employed Individuals	-0.0595	0.0454	-0.0902***	0.0297
Foreign, joint venture	0.0411	0.0454	0.0248	0.0349
Shared company	0.0721	0.0475	0.0024	0.035
Other enterprise	-0.4367**	0.1797	0.104	0.2394
Firm size: Base group= Below 8				
8-50	0.0846***	0.0321	0.0987***	0.0223
Above 50	0.1512***	0.0325	0.1235***	0.0235
Constant	1.2765***	0.1733	0.5617***	0.1073
Rho	0.6613***	0.0662	0.6227***	0.0647
Sigma	0.5107	0.0239	0.4604	0.0144
R^2	0.3614		0.3366	
N	1765		2958	

Note: Dependent variable=log of hourly wage; Provinces controlled.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 9: Observed and predicted log hourly wages

		Mean	SD
Users of network	Observed	1.6285	0.5015
	Unconditional	1.4608	0.2784
	Conditional on users of network	1.6286	0.2897
	Conditional on non-users of network	2.1705	0.3014
Non-users of network	Observed	1.7181	0.5317
	Unconditional	2.0490	0.2971
	Conditional on users of network	1.2349	0.2915
	Conditional on non-users of network	1.7176	0.3166

Table 10: Over-identifying test

	Coefficient	SE
Unmarried	-0.0103	0.0200
Non-local rural hukou	-0.0066	0.0184
Ratio of labour migrants in home village	-0.0004	0.0003
Constant	0.0237	0.0267
Adjust R^2	-0.0062	
N	4723	

Notes: Dependent variable: residual of wage equation from switching regression. All variables used in wage equation are controlled.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 11: Wage equations from Treatment effect models: ML and two-steps

	ML		Two-steps	
	Coefficient	SE	Coefficient	SE
Users of network	-0.4380***	0.0558	-0.5303***	0.1131
Rho	0.5705***		0.6631***	
Sigma	0.4662		0.4871	
Lambda	0.2660		0.3230	
N	4723		4723	

Note: Dependent variable=log of hourly wage. Personal characteristics, job characteristics and Province dummies controlled. *** $p < 0.01$

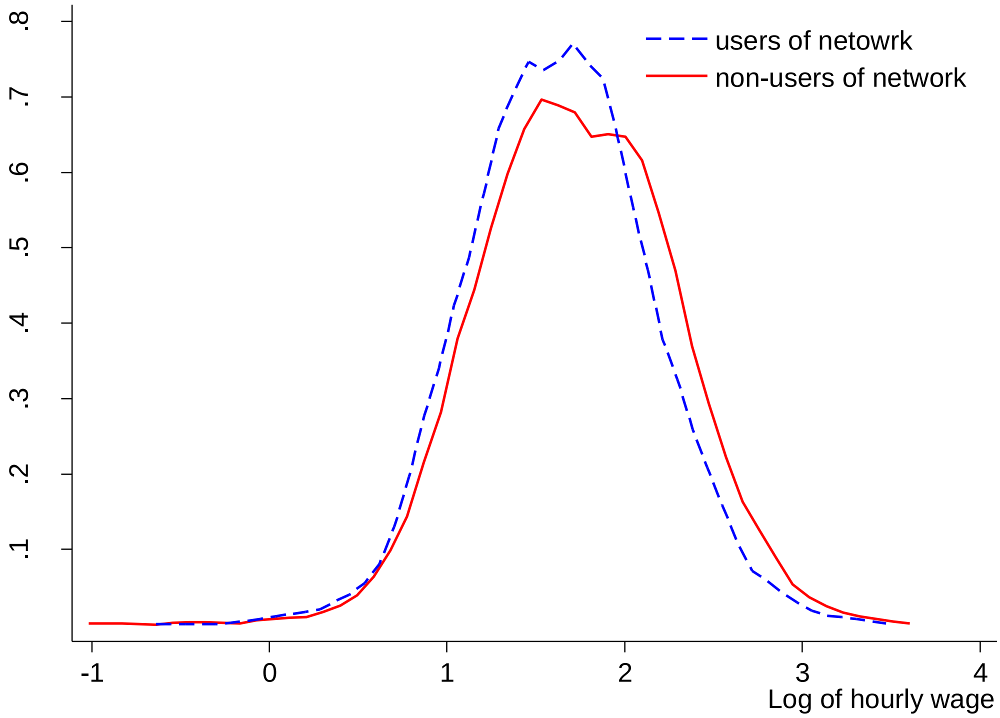

Figure 1: Kernel density graph of log of hourly wage

Appendices

Table A.1: Data descriptions (mean or frequencies)

	Non-users of network	Users of network	Total
Age	29.16	30.6	30.06
Years of migration	7.73	8.34	8.11
Years of schooling	9.46	8.68	8.97
Ratio of labour migrants in home village	56.46	59.38	58.28
Female	42.25	42.43	42.36
Male	57.75	57.57	57.64
Married	52.99	60.01	57.38
Unmarried	47.01	39.99	42.62
Local rural hukou	16.72	20.77	19.26
Non-local rural hukou	83.28	79.23	80.74
Occupation			
White collar	10.87	6.25	7.98
Blue collar	83.82	89.11	87.13
Other occupation	5.31	4.64	4.89
Industry			
Manufactory	26.71	20.95	23.1
Construction	8.44	14.41	12.18
Electricity, gas, water, IT and transportation	3.58	4.2	3.97
Commerce and trade	17.18	19.31	18.51
Restaurant and catering	20.23	19.04	19.49
Finance, estate,health, education	12.86	11.53	12.03
Services	11.00	10.56	10.73
Ownership			
State and collective	8.38	10.17	9.50
Private enterprises	40.65	43.61	42.5
Self-employed Individuals	24.39	28.73	27.1
Foreign, joint venture	15.31	9.15	11.46
Shared company	10.97	8.22	9.25
Other enterprise	0.30	0.12	0.19
Firm size			
Below 8	23.93	27.31	26.05
8-50	25.11	30.16	28.27
Above 50	50.96	42.53	45.6

Table A.2: Wage equations from Treatment effect models: ML and two-steps

	ML		Two-steps	
	Coefficient	SE	Coefficient	SE
Users of network	-0.4380***	0.0558	-0.5303***	0.1131
Age	0.0403***	0.0051	0.0401***	0.0054
Age squared	-0.0619***	0.0074	-0.0615***	0.0077
Years of migration	0.0185***	0.0039	0.0183***	0.0041
Years of migration squared	-0.0291**	0.0143	-0.0271*	0.0150
Male	0.1113***	0.0147	0.1118***	0.0153
Years of schooling	0.0328***	0.0031	0.0303***	0.0039
Occupation: Base group=White collar				
Blue collar	-0.1781***	0.0230	-0.1759***	0.0231
Other occupation	-0.1389	0.0985	-0.1389	0.0991
Industry: Base group=Manufactory				
Construction	0.1156***	0.0233	0.1131***	0.0234
Electricity, gas, water, transportation and IT	0.0250	0.0337	0.0250	0.0337
Commerce and trade	0.0082	0.0227	0.0065	0.0227
Restaurant and catering	-0.1281***	0.0213	-0.1281***	0.0213
Finance, estate, health, education	-0.0438*	0.0232	-0.0467**	0.0232
Services	-0.0665***	0.0243	-0.0657***	0.0242
Ownership: Base group=State and collective				
Private enterprises	-0.0450**	0.0221	-0.0456**	0.0221
Self-employed Individuals	-0.0770***	0.0251	-0.0762***	0.0251
Foreign, joint venture	0.0376	0.0274	0.0357	0.0274
Shared company	0.0367	0.0280	0.0342	0.0281
Other enterprise	-0.2643*	0.143	-0.2612*	0.1402
Firm size: Base group= Below 8				
8-50	0.0957***	0.0184	0.0951***	0.0184
Above 50	0.1330***	0.0191	0.1316***	0.0192
Constant	1.0695***	0.1014	1.1502***	0.1333
Rho	0.5705		0.6631	
Sigma	0.4662		0.4871	
R2	0.2660		0.3230	
N	4723		4723	

Note: Dependent variable=log of hourly wage; Provinces controlled.

* p < 0.10, ** p < 0.05, *** p < 0.01