Sociology Working Papers

Paper Number 2009-06

Network Effects among Rural-to-Urban Migrants in China

Yunsong Chen

Department of Sociology

University of Oxford

Manor Road
Oxford OX1 3UQ
www.sociology.ox.ac.uk/swp.html

Network Effects among Rural-to-Urban Migrants in China

Yunsong Chen

Department of Sociology and Nuffield College
University of Oxford

Addresses for correspondence:

Yunsong Chen, Nuffield College, New Road, Oxford, OX1 1NF, UK.
Email: yunsong.chen@sociology.ox.ac.uk

The author wishes to thank Peter Hedström, Tom Snijders, Nan Dirk de Graaf, Steve Morgan, Xiaogang Wu, Colin Mills, Scott Moser, and Mark Williams for their helpful comments and discussion. Earlier versions of this paper were presented at seminars in Oxford and China.

Network Effects among Rural-to-Urban Migrants in China

ABSTRACT

Using data from 22 provinces in China, this paper analyzes the effect of origin-based migrant networks on wages among rural-to-urban migrants, with particular attention to potential endogeneity problem. Heckman's two-stage method is used to correct for sample-selection. Natural disaster in the village of origin is used as an instrumental variable to deal with other potential estimation biases. The results presented here verify significant network effects on wages of migrants.

KEYWORDS

Network effects; Internal migration; Labour market; Endogeneity; Instrumental Variable; Heckman two-stage method

1. INTRODUCTION

Social networks are important for the analysis of migrants because they can be expected to influence migration decision-making as well as its outcomes. Although the role of social networks in migration has been widely studied (e.g., Banerjee 1983, 1984; Bauer et al. 2002; Boyd 1989; Carrington et al. 1996; Curran et. al. 2005; Curran & Rivero-Fuentes 2003; Fawcett 1989; Gurak & Caces 1992; Haug 2008; Massey et al. 1993, Massey 1987, Munshi 2003; Palloni et al 2001; Ritchey 1976; Taylor 1986; Uhlenberg 1973; to name a few), sociological analyses of how networks influence internal migration under state socialism are very scarce. In contemporary China, labour migration from rural to urban areas influences the lives of hundred millions of people. A conservative estimate from the National Bureau of Statistics of China (NBSC) indicates that the total rural-to-urban migration flow had reached 0.14 billion in 2007.

The effects of social networks on labour market outcomes among migrants are expected to be strong since they are newcomers at the destination (Borjas 1992). Recent China studies mainly focus on the network effects on migration decision-making (e.g., Hare and S Zhao 1999; Mallee 1999; Meng 1999; Zhao 1999a, 1999b, 2003; Bao et al. 2007; Lu et al. 2008). Little is known about the role of migrant networks in determining labour market outcomes. This paper thus aims to assess whether a migrant's wage at the destination is positively influenced by the size of her migrant network. The core hypothesis motivating this study is that larger networks lead to better wages. In this study, network size is measured by the number of fellow villagers who migrated to urban areas in the year before the survey year.

Notice that the measure of network size to be used is not based on administrative boundaries at the destination but on the origin villages. Migration studies (e.g., Munshi 2003 and Rainer and Siedler 2008) often use this strategy as most network relationships among immigrants at hosting areas are based on kinship, friendship, and especially belonging to a common origin (e.g., Massey et al. 1987). This strategy is even more applicable in China mainly for two reasons. First, migrant networks in China are mainly based on kinship and common village/township (see e.g., Zhao 2003; Rozelle et al. 1997). Second, most labour migrants cannot permanently settle down in cities and circular migration is the dominant form of rural-to-urban migration (e.g., Zhao 2003; Hare 1999). This makes village of origin a pivotal site facilitating the spread of job-related information among fellow villagers.

Although it is intuitive to expect that migrant networks will bring about benefits, identifying the causal effects is not easy with observational data. For example, estimation bias would occur in the presence of selection to migration on the basis of unobserved ability. Omitted village-specific or destination-specific factors would bias the estimates as well. In addition, the suspected causal relationship between wages and migrant networks can be mutual, leading to simultaneity problem. To systematically deal with sample-selection and other potential estimation biases, this paper estimates and compares an OLS model, a Heckman's two-stage model and an IV-Heckman model using natural disaster as the instrumental variable (IV). To the best of my knowledge, this is the first application of the IV-Heckit methodology in sociological analyses. The result of the empirical analysis demonstrates that the size of migrant networks significantly improves the wages of its member migrants.

The rest of the paper is structured as follows: Section 2 conducts a brief review of the literature on labour migration in China and provides the theoretical background to the analyses. Section 3 addresses the models and pays close attention to identification strategies used to correct for endogeneity biases. Section 4 describes the data and definitions of relevant variables. Section 5 presents estimation results and Section 6 proposes interpretations. Finally Section 7 concludes.

2. BACKGROUND

Large scale internal labour migration began in the late 1970s when China launched its reform and opening-up policy. The rapid economic growth from 1980s fuelled the transition of the labour force out of agriculture. In the new millennium, the annual rural-to-urban migration flow (out of township) increased from around 0.09 to 0.14 billion. The magnitude and rapid growth of internal migration from 2000 to 2007 are illustrated in Figure1 [1].

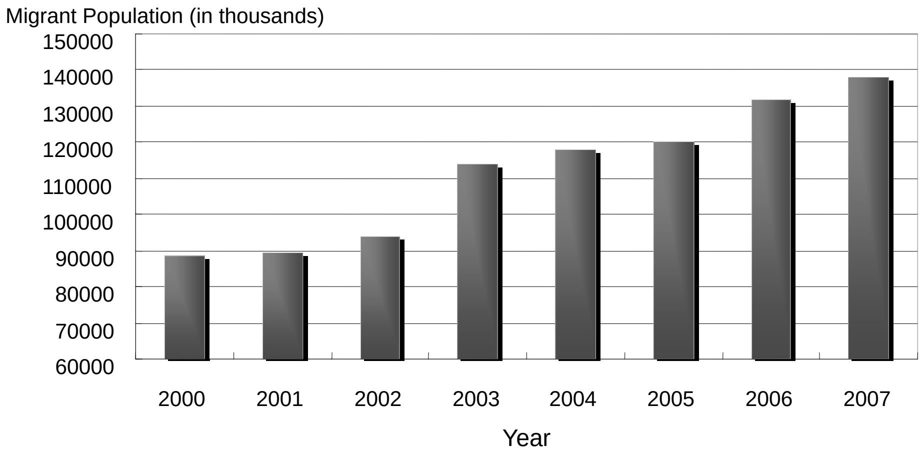

Figure. 1. Rural-to-urban migration flow in China by year

2.1 Determinants of Rural-to-Urban Migration

The prime determinants of migration decision can be classified into two types, namely push and pull factors. In particular, the expected wage gap between sending and hosting areas is the major pull factor, while surplus rural labour is often viewed as the major push factor (see e.g. Todaro 1969; Harris and Todaro 1970; Robinson and Tomes 1982). The income gap between rural and urban China after the reform and opening-up had been narrowed in the early 1980s and then widened again since 1985 onwards (Zhao 2003; Carter 1997). In 2007, the income ratio between rural and urban areas reached 3.33, which is the highest since the late 1970s [2] . Rural-to-urban migration in China is also driven by regional income gaps. This is reflected in the facts that by 2004 more than 53 per cent of rural-to-urban migration was interprovincial and 30 per cent of rural migrants went to three coastal provinces, namely Guangdong, Zhejiang and Jiangsu (Liang 2006, 2001).

The surplus labour in rural China is argued to be one of the push factors (Zhao 1999a and 1999b; Chen et. al 2008). In addition, a sizable literature on determinants of internal migration in China finds that a set of individual characteristics (e.g., health, age, gender), household and village level attributes (e.g., shortage of farmland, local rural taxation policy, distances to cities) matter for migration as well (see e.g. Jalan and Ravallion 2000; Zhao 2003; Cai 1996; Hare, 1999; see also Z Zhao2003 for a review). Importantly, many studies show that migration is selective on age and gender. Yet it is not clear how important self-selection based on unobserved factors such as earnings ability may be.

2.2 Village-Based Migrant Networks and Job Searching

Whether a market economy or a transitional economy under state socialism, it has long been recognized that many persons find jobs through their contacts or relatives (Granovetter 1974; Montgomery 1991). As for international migrants, origin-based social networks provide job-search assistance by relaying useful job leads and reducing searching costs (Massey et al. 1993; Massey 1987; Banerjee 1983, 1984; Taylor 1986, etc). This is also the case for internal migration in China. Especially, migrants in China have limited ties with the urban communities and little access to institutional support at the destination so they have to heavily rely on their origin-based networks to find jobs (e.g., Solinger 1999; Zhao 2003; Rozelle et al. 1997). Surveys show that 70 per cent of rural-to-urban migrants found jobs though the network of village-based friends or relatives (Meng 2000).

Like migrants elsewhere (e.g., Morris 2002), rural-to-urban migrants in China often become part of the underclass in destination cities. They are mainly concentrated in low-skill construction, manufacturing, and service sectors in which "chain employment" is often preferred by employers [3]. Getting jobs through referrals of earlier migrants has led to close relationships between certain occupations, hosting areas and sending areas. In general, migrants from a specific village tend to be concentrated in terms of geography and occupational types. This pattern has been also reported in international migration studies (see e.g. Aldrich and Waldinger 1990).

2.3 Temporary Return and Circular Migration

With international migration, individuals typically move to the hosting country to save up to buy a house or to invest in a business (Massey et al. 1987). Having achieved these "goals", many tend to return. In the context of internal migration in contemporary China, Murphy (1999) finds that on average one third of migrants go back to their native places. However, this type of return behaviour is temporary as most rural households regard migration as a short-term arrangement (Zhao 1999). Permanent returnees are very few in China because migrants tend to move out again after returning (Hare 1999; Zhao 2003).

Why do migrants choose not to permanently settle down in a given city given the tremendous rural-urban income gap? The first reason is the presence of the Household Registration System inherited from the planned-economy era. Despite migrants being *de facto* urban residents, there are very few channels (e.g. being formally recruited by state-owned enterprises or military demobilization) by which migrants could transfer from "agricultural" to "non-agricultural" residential status (Wu and Treiman 2007). A second reason is the lack of incentive of giving up farmland. Since farmland cannot be sold and the process for renting is quite complicated, peasants are reluctant to give up farmland (Roberts 2000).

Since permanent migration can be extremely difficult, the dominant pattern of internal migration in China is thus characterised by circular migration and temporary return. According to Hare (1999), labour migrants in China tend to have little attachment to destination cities and return home frequently during a year. An average migrant is reported to return home two to three times annually. Surveys also show that the majority of migrant workers spend less than 9 months in hosting cities (Zhao 1999). On balance, this pattern increases the role of the village of origin, making it serve as an intermediary for job leads to be transmitted among villagers at different locations. It is thus of special interest to assess the role that a village-based network has on labour market outcomes among internal migrants in China. However, so far little attention has been paid to this in existing literature.

3. MODELS AND IDENTIFICATION ISSUES

The basic specification of the model includes the size of migrant network as explanatory variable, observed individual and village characteristics as controls (e.g., age, schooling years, gender, experience in non-farm jobs, per capita net income of village, and the distance between the village to the closet county seat, etc.). However, naïvely fitting an OLS model to estimate the network effects can be problematic due to the presence of endogeneity. Hence, unobserved factors (e.g., individual heterogeneity, omitted village attributes and unobserved labour market attributes at the destination) also must be taken into account.

3.1 Sources of Endogeneity

To consistently estimate an OLS model, the explanatory variable of interest, the network size, should be uncorrelated with village-specific unobservables and individual heterogeneity. This assumption, however, is very likely to be violated due to various sources of endogeneity. This study is mainly concerned with five potential sources of endogeneity bias. First, there might be sample-selection based on unobserved abilities/skills in migration decision-making. If persons who are more “able” are more likely to migrate, the sample this paper works with would be unrepresentative. Second, since the data contain little information on the destination cities of migrants, the omitted destination factors such as local labour market conditions may bias the estimates. Third, omitted village-level attributes may bias the estimation as well. For instance, villagers from a common village may share a certain unobserved ability or reputation. Fourth, the wages and the network size may be mutually determined, leading to the simultaneity bias. For example, high wage of fellow villagers would encourage more villagers to migrate. Fifth, measurement error normally attenuates the estimated effect down toward zero.

These problems plaguing causal inferences are not new in social science. To get around them, a series of strategies are available, provided sufficiently informative data is available (Cameron & Trivedi 2005, Angrist and Krueger 1991,1999 Morgan & Winship 2007; Morgan and Harding 2006). In the sociological literature, Mouw (2006) is among the first to place emphasis on model identification issues in estimating causal effects of social capital or network resources. The approaches that can be used for alleviating these problems include fixed effects model, IV, propensity score matching, and experiments or quasi-experiments, to name a few. In this paper I combine Heckman’s two-stage approach and IV approach together to systematically deal with the endogeneity problem.

3.2 Heckman Two-stage Approach

This paper uses the Heckman two-stage approach to address the bias resting on the unrepresentative sample (Heckman 1979, Heckman and Robb 1985). The procedure of Heckit estimation is as follows: in the first step, using all observations (in this study the whole sample including both migrants and non-migrants), a probit or logit selection model (in this study the model of the migration decision) is estimated. Then the inverse Mills ratio is computed for each observation. The second step is to use the

selected sample (in this study the migrants) and fit the substantive (wage) model where the inverse Mills ratio is also included as a covariate to control for the selection factor.
To practically distinguish the sample selection from a misspecified model, however, at least one variable that does not appear in the wage model is required to be included into the selection model as an exclusion restriction (Wooldridge 2002, 2006; Cameron and Trivedi 2005). In this study, the size of household labour force (i.e., the number of household members who are aged between 16 to 60) is used as the exclusion restriction. That is, we assume that the size of household labour force influences the household decision to send its members to urban areas, while it does not directly affect the wage at the destination (see how household labour force positively affects migration in Chen et. al 2008 and Zhao 2003)

3.2 Natural Disaster as an Instrumental Variable

Although Heckman's two-stage method is useful for correcting for sample selection bias, it cannot deal with other sources of endogeneity such as omitted destination factors or village-specific factors. When using cross-sectional data, what is needed is an IV that influences the wage only by affecting the size of village-based migrant network, and which is also uncorrelated with the error term. This paper exploits the richness of data from 2002 Chinese Household Income Project Survey (hereafter CHIPS2002) and uses the intensity of any natural disaster in the origin-village as an IV. Under certain assumptions, the natural disaster can be seen as an exogenous shock to the size of outflow migrants from the village because the occurrence and destructive power of natural disaster in a certain village is random.

Although there are limited scholarly explorations on the association between migration behaviour and the environment, the close relationship is intuitive (see Hunter 2005 for a review). To make the idea clear, consider two villages of which village-A is exposed to a natural hazard while village-B not. It is easy to see that the natural disaster would decrease the expected agricultural output of the households exposed to it, and thereby encourage members of these households in village-A to migrate. Consequently, being hit by natural hazards or not matters for the outflow migration from a given village. This could be generalized to a proposition which states that the higher the intensity of a natural disaster is, the more villagers are likely to move to urban areas during the following months.

Natural hazards are common in China. Around 70 per cent of rural households are hit by natural hazards annually (Wang 2007) and the relationship between natural disasters and rural household income is very close (Gong and Zhang 2008). Importantly, in 2002, major natural disasters occurred in the first half of the year in China. From January to March 2002, drought hit 24 provinces and deprived 15.9 million rural residents of drinking-water. From April to May, strong rainstorms decimated the middle-and-lower reaches of Yangtze River [4]. Consequently, one has good reason to conjecture that the number of village migrants in 2002 should be correlated with the destructive power of the natural hazards that affected the village.

Some kinds of natural disasters might be regionally specific (e.g. earthquakes are more likely to affect some regions in China than others). And, the destructive power of them in terms of agricultural output may be different (e.g., floods influence agriculture more than earthquakes). The occurrence and intensity of natural hazards in a large region (e.g., southeast China, northwest China, or a province) are thus not random in the long term. However, if region dummies or provincial dummies can be controlled for, natural disasters in a very small tract such as a village in a given year can be comfortably seen as random and thus uncorrelated with any individual heterogeneity and omitted village-specific factors.

To be even more cautious about the exogeneity, the working assumptions are set up in such a way as to make the IV estimates as reasonable and cautious as possible: (i) Natural disaster in the origin-village is assumed to not affect any labour market conditions at the destination. (ii) Omitted village level factors are explicitly assumed to not be influenced by natural disaster within one year. (iii) Unobserved individual heterogeneity (e.g., earnings ability, propensity to migration, desire for a richer life, reservation wage), is assumed to be uncorrelated with the intensity of any natural disaster. This is important especially because natural disaster may lower the reservation wage of a potential migrant. Given the huge income gaps between rural and urban China, however, even the lowest wage levels in cities are still attractive to villagers. This makes the reservation wage argument inapplicable in this setting (see also Munshi 2003) .

3.3 OLS, Heckit and IV Models

Building on the preceding discussion about model choice and identification strategies, three different types of wage models are to be estimated, namely, the OLS, Heckit and IV-Heckit models.

First, the OLS model can be expressed by

$$W_{ig} = \beta_0 + \beta_1 S_g + \beta_2 X_{ig} + \beta_3 V_g + (U_g + e_{ig}) \quad (1)$$

where W_{ig} is the wage income of the migrant, and i indexes the individual and g indexes the village. S_g denotes the size of the migrant network, X_i denotes a vector of individual factors, and V_g is a vector of observed village level factors. U_g is the combined effect of various omitted village level factors and ϵ_{ig} is the combined effects of individual unobservables. Thus, S_g is the major explanatory variable of interest and β_1 denotes the network effect. Notice that one of the most important premises for obtaining unbiased OLS estimates of β_1 is that Cov [S_g, ϵ_{ig}] =0 and Cov [S_g, U_g] =0. However, these two assumptions do not hold in the context of this study because of the potential presence of sample selection and other conventional endogeneity problem. As a result, alternative models are needed.

The Heckit model is:

$$W_{ig} = \beta_0 + \beta_1 S_g + \beta_2 X_{ig} + \beta_3 V_g + \beta_4 \hat{P}_{ig} + (U_g + e_{ig}) \quad (2)$$

where $\hat{P}_{ig}$ is the inverse Mill ratio, a transformation of the predicted individual probability of migration calculated from the following selection equation (probit estimation):

$$P_{ig} = \gamma_0 + \gamma_1 F_{ig} + \gamma_2 S_g + \gamma_3 X_{ig} + \gamma_4 V_g + \mu_{ig} \qquad (3)$$

where F_{ig} is the household labour force which is not included in equation (2). As noted before, the explanatory variables in equation (2) should be a strict subset of those of the equation (3).

Finally, based on the Heckit model, the IV model can be written as the following equation system：

$$W_{ig} = \beta_0 + \beta_1 S_g + \beta_2 X_{ig} + \beta_3 V_g + \beta_4 \hat{P}_{ig} + (U_g + \epsilon_{ig}) \qquad (4)$$

$$P_{ig} = \gamma_0 + \gamma_1 F_{ig} + \gamma_2 N_g + \gamma_3 X_{ig} + \gamma_4 V_g + \mu_{ig} \qquad (5)$$

$$S_g = \gamma_0 + \gamma_1 N_g + \gamma_2 X_{ig} + \gamma_3 V_g + \gamma_4 \hat{P}_{ig} + \mu_g \qquad (6)$$

where N_g is the intensity of the natural disaster and μ_g is the random error term in the equation denoting how the network size is influenced. Notice that in equation (5) the endogenous explanatory variable S_g is not included but the exogenous instrumental variable N_g is included. The specification of the equation systems can be referred to Wooldridge (2002) .

In the first step, we estimate the selection equation (5) by Probit and then calculate $\hat{P}_{ig}$ (equation 5).

We then used N_g as an instrument for S_g. The process is: given Cov[U_g, N_g] = 0, Cov[ϵ_{ig}, N_g] = 0, and Cov [N_g , S_g] $\neq$ 0, we regress S_g on N_g and other exogenous variables in the first-stage regression (equation 6) and get $\hat{S}_g = \hat{\gamma_0} + \hat{\gamma_1} N_g + \hat{\gamma_2} X_{ig} + \hat{\gamma_3} V_g + \hat{\gamma_4} \hat{P}_{ig}$. Then in the second-stage regression (equation 4) we use $\hat{S}_g$ instead of S_g to get consistent estimates of β_1 since $\hat{S}_g$ must not be correlated with U_g and ϵ_{ig} as long as the IV is valid. Notice that the interpretation for IV estimate can be complicated if the assumption that network effects are constant among migrants is relaxed. This will be discussed in detail below.

4. DATA AND VARIABLES

The CHIPS 2002 was jointly conducted by the Rural Survey Group of National Bureau of Statistics of China (NBSC) and the Institute of Economics of the Chinese Academy of Social Science (CASS). The questionnaires were collected in February 2003 during the Chinese Lunar New Year. The advantage of doing so is that most outward migrants returned home to reunite with their family members. The survey thus captures the information of most members of rural households. The CHIPS 2002 gathered detailed information on demographic, household and village characteristics of the rural respondents. The data used in this study are from a sample of 37,969 individuals from 9,200 households, 961 administrative villages, 121 counties and 22 provinces[5]. The surveyed areas are filled with grey in Figure 2. To focus on sample selection in migration decision and conventional omitted variable problems, here we assume that returning home in the Lunar New Year is not selective.

Figure 2. Surveyed Provinces in CHIPS2002

Different criteria in defining the migrant population have been used in different national surveys in China (Liang 2006). The CHIPS 2002 asked the respondent to report the days away from their family in 2002 as well as the type of destination place. The precise questions read: “How many days did you live outside this household in 2002?” and “Place of non-agricultural waged work: 1) within native village; 2) out of native village, within native township; 3) out of native township, within native county; 4) out of native county, within native province; 5) out of native province”. Exploiting this information, this paper sets the duration criterion as 180 days and destination criterion as being away from the township where the migrants were born [6]. This is consistent with the criteria of NBSC and previous migration studies (e.g., Zhao 2003; Chen et al. 2008). Thus, migrants are defined as individuals (none-student) with the "agricultural" *hukou* status who lived outside of their native-township for more than 180 days in 2002.

This definition seems a bit arbitrary, but other definitions of migrant can and will be employed in robustness tests to examine whether the empirical results are sensitive to alternative definitions. In addition, all long-term out-of-township students, respondents younger than 16, and respondents older than 60 are dropped from the sample. Finally, due to some other variables of interest with missing values, the effective sample size ended up comprising 20,040 individuals from 8,308 households, 919 villages and 121 counties. Among them, there were 2,382 migrants from 1,800 households, 567 villages and 112 counties.

The dependent variable is the daily wage of migrants in 2002 at the destination. This is calculated by the reported annual wage divided by the days living at the destination. Due to the malpractice of local government and loopholes in the legal system, the problem of wage arrears among migrant workers has been severe in China, triggering social unrest and violence (see e.g. Li 2008; Hannan 2008). This explains why one can find in Table 1 that the minimum waged income is actually zero. In the sample, 280 out of 2,382 migrants went home empty-handed in 2002. As Table 1 shows, the average daily wage of the migrants is 12.5 RMB (*yuan*) and the maximum wage is 150 RMB. Excluding those without pay in 2002, the average daily wage is 14.2 RMB which is in general consistent with some previous studies (Li 2008; Hannan 2008) [7].

Table 1： Selected Descriptive Statistics

	Migrants (N=2382)				Non-Migrants (N=17658)			
	Mean	**Min**	**Max**	**S.D.**	**Mean**	**Min**	**Max**	**S.D.**
Daily Wage 2002 (¥)	12.54	0	150	11.98	--	--	--	--
Age	26.98	16	60	8.25	38.23	16	60	11.85
Age-Squared	796	256	3,600	532	1,602	256	3,600	903
Experience	4.00	0	38	3.58	2.00	0	43	4.75
Experience-Squared	28.74	0	1,444	71.35	26.61	0	1,849	103.70
Years of Schooling	8.14	0	16	2.19	7.02	0	16	2.67
Household Labour Force	4.80	2	11	1.38	4.44	1	2	1.35
Village Migrant Network Size	252	0	1,530	251	153	0	1,530	189
Per Cap. Village Income 98	1,860	0	13,980	1,026	2,079	0	13,980	1183
Village Population 2002	2081	186	8,815	1379	1,876	186	8,815	1233
Village Arable Land 2002	3,436	0	15,883	2808	3,566	0	25,463	2879
Distance from County (km)	27.07	1	140	20.61	24.27	.5	160	21.l1
Intensity of Natural Disaster	.11	0	.85	.13	0.12	0	0.9	.16

	Migrants Freq.	Migrants per cent	Non-Migrants Freq.	Non-Migrants per cent
Male	1,493	62.67	9,233	52.29
Female	889	37.32	8,425	47.71
Married	1,038	43.82	14,362	81.59
Unmarried/Single	1,331	56.18	3,240	18.41
Suburb of Large/Middle City	61	2.56	1,372	7.77
None-Suburb	2,321	97.44	16,286	92.23

The size of the migrant network is the explanatory variable of central interest. Employing the origin community as a "well-defined and well-established" social unit to construct the measure of migrant network does have its advantages (Munshi 2003:551). Aside from capturing the feature of migrant networks (i.e, highly village-based social ties), it offers the opportunity to instrument the explanatory variable using natural disaster. The measure of network size is obtained from the responses of the village heads to a question posed in the survey. Specifically, all the village heads were asked about the number of outflow migrants in 2002 from her village. The precise question is "Approximate total number of out-migrants working out of township for more than six months".

Notice that although the head of a village in China is elected by villagers elder than 16, the township authority plays an important role by naming the candidates and inspecting the election process. The village head is thus required to report key statistics of the village regularly (e.g., migration, fertility, and agriculture products) to the township government. More importantly, there is little motivation for village heads to over or under report the number of migrants to the surveyors from academic institutions. These facts suggest that the network size of migrants reported by village heads is measured validly.

The individual level controls include the migrants' age, age square, gender, schooling years, experience, experience square, type of economic sector, and region dummies. Note that since it is the off-farm experience that matters for the waged income at the destination, the experience variable is measured by how many years migrants have worked in cities (i.e., 2003 minus the starting year of working in cities). Although not reported in Table 1, rural-to-urban migrants are engaged in 18 types of occupations. Among them, more than 60 per cent of migrants are engaged in the "industry" and "construction" sectors. Also not presented in Table 1 is the distribution of migrants across 22 sending provinces. As noted before, region dummies should be included in the model to assure that the natural disaster can be seen as random. According to different natural/climate conditions and social-economic development of villages of origin, the 22 provinces are classified into 5 regions, namely Metropolis Suburb (Beijing and Shanghai as two municipalities), Eastern China, Southern China, Northern China, and Inner (Western and Middle) China. The village-level controls include the per capita net income (including both migrants and non-migrants) in 1998, the distances to the closest county seat, and a dummy denoting whether the village is a suburb of a large/middle city or not. Including the lagged mean income of the village is an attempt to proxy for unobserved village-level earning ability. The distance to the closest county seat and being a suburb of a large/middle city both proxy for access to beneficial job leads.

The intensity of the natural disaster that serves as the IV is measured by the percentage reduction in agricultural output in the village. For each village surveyed by CHIPS 2002, the village head was asked to answer the question "Was there any natural disaster during the year of 2002" and "If yes, what was the percentage reduction in agricultural output as compared to an ordinary year?" For those villages without the presence of any natural hazard in 2002, the variable is coded as zero representing that there is no loss induced by the natural disaster. The village heads who could not provide accurate figures were asked to provide rough estimates for the percentage reduction in agricultural output. As can be seen in Table 1, the average percentage reduction in agricultural output among migrants is around 11 per cent with a maximum percentage of 85 per cent.

5. ESTIMATION RESULTS

The analysis begins with a standard OLS regression. However, as discussed above, the OLS estimator may yield biased estimates as the network size may be endogenous. Since the goal of this analysis is causal inference, or identification of the mechanism of network effects, rather than to merely describe the statistical association, it is necessary to seriously deal with the potential endogeneity problem [8]. To this end, the Heckman two-stage approach is employed to address sample selection problem. Since there may still be simultaneity bias, conventional omitted variable bias and measurement error in addition to sample selection, ultimate results are obtained from an IV-Heckit model explicitly taking into account both sample selection and the other endogeneity problems. The OLS, Heckit and IV-Heckit models, the estimated coefficients and p-values are presented in Table 2. When fitting models, the wage and network size variables appear in logarithm form given that the skewness statistics for them (4.46 and 2.52 indicating considerable right-handed skewness).

The first column of estimation results in Table 2 contains the OLS estimates [9]. The coefficient of the network size variable is around 0.13 with an extremely small p-value (less than 0.001), suggesting that migrants who came from a village with 8 per cent larger migrant network (outflow migrants) are associated with 1 per cent higher wages. While the age and age square variables both significantly influence wages at the 0.05 significance level, the directions of the two effects are different. This implies a diminishing marginal effect of the age. Experience of working in cities is also an important determinant of wages, yet its squared value is not statistically significant, suggesting no diminishing return to experiences.

Also as expected, years of schooling do not seem to be related to the wages of the migrant workers. This could be due to fact that the occupations that migrants take in contemporary China are extensive but mainly low-skill and the average educational requirement is quite low. What is unexpected is the role of gender. According to the empirical result, the wage difference between men and women is not significant. This is probably because of the limited occupation choices as well as the ceiling effects in terms of promotion among rural-to-urban migrants. On the one hand, they are mainly concentrated in low-income jobs in which the gender difference in wage is marginal all else being equal; on the other hand, the opportunities for them to be promoted within workplaces is very limited so that male migrants do not have more advantages than women. Finally, in general the village characteristics do not affect wages significantly.

The second column presents the Heckit estimates. When selection in migration on unobserved ability is explicitly taken into account, the network effect increases from 0.13 to 0.23, and remains highly significant with a p-value of less than 0.001. The partial effect of the latent selection factor (i.e., the predicted probability of migration which proxies the unobserved factors influencing both migration and incomes), is around 0.64 and significant at the 5 per cent level. These findings strongly suggest that: 1) there is considerable selection based on unobserved individual ability when people make migration decisions; and 2) more able rural residents tend to migrate than the less able ones, everything else being equal. That is, the Heckit estimate of network effects reveals that the OLS estimate may have largely reflected the network effects among more able people who benefited less from the origin-based networks. Of interest is why less-able migrants benefit more from village-based migrant networks than more able migrants? This may be explained by the fact that migrants of higher earning-ability are also more able to initiate social ties with urban communities at the destination. Thus their job-searching rely less on village-based networks since the new ties built at the destination generate more valuable job leads.

Table 2： OLS, Heckit and IV Models of Migrants' Wage

Logged Wages as the Dependent Variable (N=2382)

Independents Variables	OLS	Heckit	IV-Heckit
Logged Network Size	.131***	.228***	.584**
	(0.000)	(0.000)	(0.039)
Age	.0688***	.0499**	.0502**
	(0.001)	(0.014)	(0.024)
Age-Squared	-.00101***	-.00109***	-.00109***
	(0.002)	(0.000)	(0.000)
Gender	.0187	.0174	.0176
	(0.695)	(0.740)	(0.750)
Years of Schooling	.0124	.0191*	.0172
	(0.319)	(0.098)	(0.164)
Experience	.0667***	.105***	.0767**
	(0.001)	(0.000)	(0.017)
Experience-Squared	-.00185	-.00387***	-.00303**
	(0.052)	(0.001)	(0.039)
Per Cap. Village Income 1998	.0365	-.0197	-.0845
	(0.316)	(0.606)	(0.165)
Distance to Closest City	.00497**	.00782***	.00527*
	(0.005)	(0.000)	(0.076)
Suburb of Big City	-.0403	.271	-.00428
	(0.833)	(0.189)	(0.989)
Occupation Dummies	YES**	YES**	YES**
Region Dummies	YES	YES	YES
Constant	4.82***	5.48***	-3.26**
	(0.000)	(0.000)	(0.013)
Latent Selection Factor		.64**	.63**
		(0.033)	(0.047)
First stage F-statistic on IV			19.27
Hausman test of endogeneity			p=0.046

Note: i) Standard errors are in parentheses and robust to heteroscedasticy;

ii) * p < 0.1 ** p < 0.05, *** p < 0.01

Several important results are found in the third column which contains the IV-Heckit estimates. First of all, the quality of the IV is good since it comfortably passed underidentification and weak-identification test. As is shown at the bottom of the third column, the F-statistic for significance of the IV in first-stage regression of the endogenous network size is 19.27 and significant at 0.05 alpha level. In light of Stock and Watson's (1997) rule of thumb that the first-stage F-statistic should be at least 10 for a single endogenous regressor, this implies a significant statistical association between the natural disaster and village migration. It is noteworthy that the F-statistic is also larger than the relevant Stock and Yogo's (2003) weak IV critical value for (16.38 for one endogenous regressor and one IV), suggesting that the relevance of IV is sufficiently strong to reject the weak IV hypothesis with less than 10 per cent relative bias toleration. This makes us fairly confident in the IV estimates. More importantly, the Hausman endogeneity test indicates that we can reject the null hypothesis that there is no systematic difference between the estimates from the OLS and the IV model. Since the IV estimator always yields consistent estimates given a valid IV, coefficients estimated from it should be more meaningful. This further suggests that the size of the migrant network is indeed endogenous.

Compared to the Heckit estimate, the effect of network size obtained from the IV model is substantially higher. It increases from less than 0.23 to around 0.58. The standard error of the IV estimate increases as well and the p-value increases from less than 0.001 to 0.039, which indicates that the effect remains significant given the 5 per cent significant level. As for other controls and the latent selection factor, no major changes are found in the partial effect. Consider two villages sending migrants to the same city. All else being equal, migrants who came from villages with 2 per cent larger migrant network receive around 1 percent higher wages. However, an issue arises that the larger IV estimate of the network effects looks counter-intuitive. This is because one would expect that correcting for the remaining part of estimation bias would lead to a mitigated IV estimate as compared to the Heckit estimate. This will be discussed in the following section.

Before proceeding to the discussion of the IV estimate, the results should be examined to see whether they are qualitatively robust with respect to changes in the definition of migration and differently truncated samples to deal with wage outliers. To this end, migrants living out-of-native-village more than 180 days, out-of-native-county more than 180 days, out-of-native-township more than 90 days and out-of—native-township more than 240 days are examined respectively. In addition, considering the potential outliers in terms of wages, we tried deleting observations in top 99 per cent, top 95 per cent, bottom 99 per cent, bottom95 per cent quantiles from of the sample. Also, the sample without those who report zero-wage is also used to see whether the analysis is sensitive to zero-wage.

As can be seen in Table 3, all of the estimates based on different definitions of migration (from definition 1 to definition 4) show identical patterns compared to the estimates under "definition 0" which are those repeated in Table 2, suggesting that the results are not sensitive to the definition of migrants. Likewise, using differently truncated samples does not lead to any considerable changes in estimates compared to the results from the "full sample" which is also repeated in Table 2. All of these results suggest that the estimates in this study are robust.

Table 3: Robustness Tests

Different Operational Definitions of Migrants

	Definition0	Definition1	Definition2	Definition3	Definition4
Destination	**OOT**	**OOV**	**OOC**	**OOT**	**OOT**
Duration	**>180 days**	**>180 days**	**>180 days**	**>90 days**	**>240 days**
OLS estimates	.131	.151	.127	.096	.139
p-value	0.000	0.000	0.000	0.002	0.000
Heckit stimates	.228	.214	.250	.154	.261
p-value	0.000	0.000	0.000	0.000	0.000
IV estimates	.584	.773	.619	1.058	.829
p-value	0.039	0.024	0.021	0.022	0.032
N	2382	2486	2135	2800	1975

Notes: OOT=Out of township OOV=Out of village OOC=Out of county

Different Working Samples (in terms of wages)

	Full samlpe	Top 99 per cent deleted	Top 95 per cent deleted	Zero-wage deleted
OLS estimates	.131	.137	.143	.0722
p-value	0.000	0.000	0.000	.0007
Heckit estimates	.228	.194	.186	.2516
p-value	0.000	0.000	0.000	0.000
IV estimates	.584	.605	.731	.468
p-value	0.039	0.033	0.014	0.035
N	2382	2359	2254	2102

6. INTERPRETATION OF THE IV ESTIMATE

The IV-Heckit estimate suggests that network effects obtained from conventional Heckit model are downwardly biased. One possible reason for this is that the omitted variable is positively correlated with network size yet negatively correlated with wages, holding the latent selection factor and other observed factors constant. Notice that to this point the effect of migrant networks has been assumed to be constant across individuals and villages. However, this assumption may be too strong in the context of rural-to-urban migration. It thus seems reasonable to interpret the IV estimate within the framework of heterogeneous network effects.

If natural disasters influence migration mainly among villages/individuals benefiting from more extensive networks, the IV estimate would be higher than the Heckit estimate. This is simply because the Heckit estimate applies to all villages/individuals while the IV estimate mainly applies to the subgroup of villages/individuals more affected by natural disasters. In the language of econometrics, with the presence of casual effects heterogeneity, the IV estimate can be interpreted as a Local Average Treatment Effects (LATE) (Imbens and Angrist 1994; Angrist et al 1996)[10]. For example, Cards (1995) uses college proximity as an IV to identify the returns to schooling. He reports larger IV estimate than conventional OLS estimate, and accounts for it by arguing that the college proximity affects schooling mainly among those children who have poorly educated parents and thereby benefit more from education.

If the IV estimate is to be interpreted as a class of LATE, we must raise the question about the mechanism that explains how natural disasters influence migration and why networks effects differ across villages/individuals. One possible mechanism is that that less able people (in terms of earning ability at the destination) are more responsive to natural disasters since they have relatively lower ability to compensate for losses due to natural disasters. That is, villagers of lower earning ability are more likely to be "pulled" out from the villages by natural hazards. If this is the case, the IV estimate can be interpreted as a weighted average network effect and the weight for less able migrants is relatively higher. Since we know less able migrants benefit more from the origin-based networks, the IV estimate mainly captures the network effects among them and it would then be higher than the Heckit estimate[11].

7. CONCLUDING REMARKS

The rural-to-urban migration in China since the late 1970s represents the largest labour flow ever observed in the world. Despite the proliferation of research seeking to identify the mechanisms and measure the magnitude of internal migration, little emphasis has been placed on probing the direct causal effects of migrant networks on labour market outcomes at the destination. This paper explores the causal effects of the size of migrant networks on wages among migrants in China. It complements recent research on the labour market effects of migrant networks in market economies such as the U.S. and Germany (e.g., Munshi 2003; Rainer and Siedler 2008). The major innovation of this paper is the combination of Heckman's two-stage method and IV approach. As far as the author knows this is the first IV-Heckit application in a sociological analysis of network effects.

Controlling for the unobserved factors influencing migration decision, identification is achieved through instrumenting the network size by the intensity of natural disasters occurring in the sending-villages of the migrants. The empirical results show that the size of the migrant network significantly improves the wages of migrants. This is consistent with findings reported elsewhere and implies that the mechanism of job-search among migrants in the context of state socialism is similar to that in other contexts. As compared to the OLS and conventional Heckit estimates, the IV-Heckit estimates are larger. These findings suggest the existence of casual effects of network size among at least a subgroup of rural-to-urban migrants. In addition, estimates obtained from Heckit model are higher than those from OLS model, and the wage effects of selection factors are positive. This suggests the presence of positive selection in terms of unobservables (e.g., earning ability) in migration.

One potential criticism of this study is that the validity of the IV approach cannot be statistically tested for. This might be one of the reasons why sociologists rarely use the IV approach. However, as Morgan (2002) puts it when he calls for more attention to be paid to the IV approach, "within the new literature on counterfactual causality, instrumental variables estimators have received substantial attention from both statisticians and econometricians … even fallible IVs can be useful for examining patterns of causal effect heterogeneity." Surely the results reported in this paper should not be regarded as definitive. Yet the results and the identification strategy improve our understanding on the mechanism though which cohesive migrant networks influences wages of migrants.

Notes

1. According to National Bureau of Statistics of China (NBSC) the rural-to-urban migration is defined as working outside of township more than one month in a year. Given the different definitions of migrant population used in surveys and censuses, estimation of China's migrant population varies. However, the huge population and rapid growth is a stylized fact.
2. The data is cited from the 2008 state message of State Council of China titled as "Spurring steady increase in rural incomes".
3. See an introduction on chain employment in Grieco (1987).
4. The data are obtained from the 2002 Report on the State of Environment in China by Ministry of Environmental Protection of the PRC.
5. The sampling-frame for CHIPS2002 is a subsample of the official rural household survey conducted by the NBSC. The administrative village, as opposed to natural village, refers to a village which is legally recognized and numbered accordingly by the government. That is, one administrative village may consist of more than one natural village.
6. Since 1997 six-month duration has become the standard criteria for defining migrants and was used in the recent Chinese Population Census by the NBSC (Liang 2006).
7. The conservative estimation of the average monthly wage of migrant workers in 2002 is around 500RMB and optimistic estimation is around 700 RMB.
8. See the definition of mechanism in Hedström (2005) and the association between causality, mechanisms and counterfactual framework in Morgan and Winship (2007).
9. Since villagers are "clustered" in villages, the observations of this study may not be independent. To correct for the intra-village correlation, this study calculates clustered error terms when estimating OLS model.
10. Importantly, interpreting IV estimates under the LATE framework should be on the condition that the relationship between the IV and the explanatory variable is of monotonicity. That is, the natural disaster should influence the village migration in one direction.
11. An alternative explanation can be that the distribution of marginal land (e.g. the lands get flooded most easily within a village) is correlated with individual characteristics (the author wishes to thank Colin Mills for proposing this). For example, the poorest or less able people live in the most marginal land and therefore are more likely to be affected by natural disaster. And, the poorest most benefit from migration networks. In addition, migration tradition of villages can also be a potential explanation because natural disaster would "pull" more people from the villages with weak migration tradition, and these villages may have less able villagers due to limited access to urban areas. Due to limited space, I have not included them into the paper.

References

Aldrich, Howard and Roger Waldinger. 1990. "Ethnicity and Entrepreneurship." Annual Review of Sociology. 16: 111-135.

Angrist, Joshua D., and Alan B. Krueger. 1991. "Does Compulsory School Attendance Affect Schooling and earnings?" The Quarterly Journal of Economics. 106(4): 979-1014.

Angrist, Joshua D., and Alan B. Krueger. 1999. "Empirical Strategies in Labour Economics."In Orley Ashenfelter & David Card, (eds.) Handbook of Labour Economics IIIA. Amsterdam: North-Holland.

Angrist, Joshua D., Guido W Imbens, and D.B. Rubin. 1996. "Identification of Causal Effects Using Instrumental Variables." Journal of the American Statistical Association. 91: 444-455.

Banerjee, Biswajit. 1984. "Information Flow, Expectations and Job Search: Rural-to-Urban Migration Process in India." Journal of Development Economics. 15: 239-257.

------ 1983. "Social Networks in the Migration Process: Empirical Evidence on Chain Migration in India." The Journal of Development Areas. 17: 185-196.

Bao, Shuming, Örn B. Bodvarsson, Jack W. Hou, and Yaohui Zhao. 2007. "Interprovincial Migration in China: The Effects of Investment and Migrant Networks." IZA Discussion Paper No: 2924.

Borjas, George. 1987. "Self-Selection and the Earnings of Immigrants." American Economic Review, LXXVII : 531–553.

——1992. "Ethnic Capital and Intergenerational Mobility." Quarterly Journal of Economics, CVII : 123–150.

—— 1995. "Ethnicity, Neighborhoods and Human-Capital Externalities." American Economic Review, LXXXV : 365–390.

Bauer, Thomas, Pedro T. Pereira, Michael Vogler, and Klaus F. Zimmermann. 2002. "Portuguese Migrants in the German Labour Market: Selection and Performance." International Migration Review, 36(2): 467-491

Boyd, Monica. 1989. "Family and Personal Networks in International Migration: Recent Developments and New Agendas." International Migration Review 23(3):638-670.

Cai, Fang. 1996. "An Economic Analysis for Labour Migration and Mobility." China Social Sciences Quarterly, Spring, 120-135.

Cameron, A. Colin, and Pravin K. Trivedi. 2005. Microeconometrics: Methods and Applications. Cambridge：Cambridge University Press.

Cao, Yang, and Victor Nee. 2000. "Comment: Controversies and Evidence in the Market Transition Debate." American Journal of Sociology. 105(4):1175-1189.

Card, David. 1995. "Using Geographic variation in college proximity to estimate the return to schooling.” In L.N. Christofides, E.K. Grant, and R. Swidinsky, (eds.) Aspects of Labour Market Behaviour: Essays in Honour of John Vanderkamp. Toronto: University of Toronto Press, 1995.

Carrington, William J., Enrica Detragiache, and Tara Vishwanath. 1996 “Migration with Endogenous Moving Costs.” The American Economic Review. 86(4): 909-930.

Chen, Zhao, Shiqing Jiang, Ming Lu, and Hiroshi Sato. 2008. “How do Heterogeneous Social Interactions affect the Peer Effect in Rural–Urban Migration?:Empirical Evidence from China.” LICOS Discussion Papers 22408, Centre for Institutions and Economic Performance, K.U.Leuven.

Curran, Sara R., and Rivero-Fuentes E. 2003. “Engendering Migrant Networks: The Case of Mexican Migration.” Demography 40: 289-307.

Fawcett, James T. 1989. “Networks, Linkages, and Migration Systems.” International Migration Review 23: 671-680

Gong, Qianwen and Zhang, Junbiao. 2007 “Relations between Agricultural Natural Disasters and Rural Poverty Based on Panel Analysis in Anhui Province.” Chinese Population, Resources and Enviroment:17(4).

Granovetter, Mark S. 1973. “The Strength of Weak Ties.” American Journal of Sociology. 78(6): 1360-1380.

------1974:1995. Getting a Job: A Study of Contacts and Careers. Cambridge, MA:Harvard University Press.

Grieco, Margaret. 1987. Keeping it in the family: social networks and employment chance. Tavistock Publications: London.

Gurak, Douglas T., and Fe Caces. 1992. “Migration Networks and the Shaping of International Migration Systems.” In Mary Kritz, Lin Lean Lim, and. Hama Zlotmk, (eds.) International Migration Systems: A Global Approach. Oxford: Clarendon Press: 150-176.

Hare, Denise. 1999. “ ‘Push’ versus ‘Pull’ Factors in Migration Outflows and Returns: Determinants of Migration Status and Spell Duration among China’s Rural Population.” Journal of Development Studies. 35(3): 45-72.

Hare, Denise and Shukai Zhao. 2000. “Labour Migration as a Rural Development Strategy: A View from the Migration Origin.” In West and Zhao (eds.) Rural Labour Flows in China, Institute of East Asian Studies, University of California, Berkeley.

Harris J. R., and Todaro M.P. 1970. “Migration, Unemployment and Development: A Two-Sector Analysis.” American Economic Review 60:126-142

Hannan, Kate. 2008. “ China: Migrant Workers Want 'Decent Work'.” The Copenhagen Journal of Asian Studies. 26(2).

Haug, Sonja. 2008. “Migration Networks and Migration Decision-Making.” Journal of Ethnic and Migration Studies. 34: 585-605

Heckman James. J. 1979. “Sample Selection Bias as a Specification Error.” Econometrica. 47: 153–161

Heckman James. J. and R. Robb. 1985. “Alternative Methods for Evaluating the Impact of Interventions: An Overview.” Journal of Econometrics. 30: 239–267.

Hedström. Peter. 2005. Dissecting the Social: on the Principles of Analytical Sociology. Cambridge: Cambridge University Press.

Hunter, Lori M. 2005. “Migration and Environmental Hazards.” Population and Environment 26 (4): 273–302.

Imbens, G. W., and J. D. Angrist. 1994. “Identification and Estimation of Local Average Treatment Effects.” Econometrica. 62(2): 467–475.

Li, Shi. 2008 “Rural Migrant Workers in China: Scenario, Challenges and Public Policy.” Working Paper .

Liang, Zai. 2001. “The Age of Migration in China”. Population and Development Review. 27(3): 499-524.

---- 2006. “Internal Migration in China during the Reform Era: Patterns, Policies, and Challenges.” Working Paper.

Lin, Nan. 1999. “Social Networks and Status Attainment.” Annual Review of Sociology. 25:467-487.

Mallee, Hein. 2000. “Agricultural Labour and Rural Population Mobility: Some Observations,” in West and Zhao (eds.) Rural Labour Flows in China, Institute of East Asian Studies, University of California, Berkeley.

Massey, Douglas, S., Joaquin Arango, Graeme Hugo, Ali Kouaouci, Adela Pellegrino, and J. Edward Taylor, 1993. “Theories of International Migration: A Review and Appraisal.” Population and Development Review. 19(3): 431-466.

Massey, Douglas S., Rafael Alarcon, Jorge Durand, and Humberto Gonzalez, 1987. Return to Aztlan: The Social Process of International Migration from Western Mexico. Berkeley: University of California Press.

Massey, Douglas S., 1987. “Understanding Mexican Migration to the United States,” American Journal of Sociology. 92(6): 1372-1403.

Meng, Xin. 2000, Labour Market Reform in China, Cambridge: Cambridge University Press.

Montgomery, James D. 1992. “Job Search and Network Composition: Implications of the Strength-of-Weak-Ties Hypothesis.” American Sociological Review 57: 586-96.

------1994. “Weak Ties, Employment and Inequality: An Equilibrium Analysis.” American Journal of Sociology 99(5): 1212-1236.

Morgan, Stephen L. and. David J. Harding. 2006. “Matching Estimators of Causal Effects: Prospects and Pitfalls in Theory and Practice.” Sociological Methods and Research 35:3-60.

Morgan, Stephen L. 2002. “Should Sociologists Use Instrumental Variables?” Working Paper, Cornell University

Morgan, Stephen L. and Christopher Winship. 2007. Counterfactuals and Causal Inference: Methods and Principles for Social Research. Cambridge: Cambridge University Press.

Morris, Lydia. 2002. Managing Migration: Civic Stratification and Migrants Rights, Routledge, London

Mouw, Ted. 2003. “Social Capital and Finding a Job: Do Contacts Matter?” American Sociological Review. 68(6): 868-898.

------ 2006.“Estimating the Causal Effect of Social Capital：A Review of Recent Research.”Annual Review of Sociology. 32:53-78.

Munshi, Kaivan. 2003. “Networks in the Modern Economy: Mexican Migrants in the U.S. Labour Market.” Quarterly Journal of Economics, 118(2): 549-599.

Murphy, Rachel, 1999. “Return Migration Entrepreneurs and Economic Diversification in Two Counties in South Jiangxi, China.” Journal of International Development. 11: 661-672.

Palloni, Alberto, Douglas S. Massey, Miguel Ceballos, Kristin Espinosa, and Michael Spittel. 2001. “Social Capital and International Migration: A Test Using Information on Family Networks.” The American Journal of Sociology 106: 1262-1298 .

Rainer, Helmut, and Tom Siedler. 2008. “Social networks in determining migration and labour market outcomes: Evidence from the German Reunification.” ISER Working Papers 2008-36.

Ritchey, P. Neal. 1976. “Explanations of Migration.” Annual Review of Sociology 2 : 363-404.

Roberts, Kenneth D., 2000. “Chinese Labour Migration: Insights from Mexican Undocumented Migration to the United States,” in West and Zhao (eds.) Rural Labour Flows in China, Institute of East Asian Studies, University of California, Berkeley.

Robinson, Chris, and Nigel Tomes. 1982. “Self-selection and Interprovincial Migration in Canada.” Canadian Journal of Economics 15:474-502

Rozelle, Scott, Li Guo, Minggao Shen, John Giles and Tuan Yee Low, 1997. “Poverty, Networks, Institution, or Education: Testing among Competing Hypotheses on the Determinants of Migration in China,” Paper presented at the 1997 Annual Meetings of the Association for Asian Studies, Chicago, Illinois, March 13-16.

Solinger, Dorothy J. 1999. “Citizenship Issues in China's Internal Migration : Comparisons with Germany and Japan.” Political Science Quarterly. 114(3)455-478.

Staiger, D., and J. H. Stock. 1997. “Instrumental Variable Regression with Weak Instruments,” Econometrica, 65(3): 557–586.

Stock, James H., and Mark W. Watson. 1997. Introduction to Econometrics (2nd edition). Addison Wesley

Stock, James H., and Motohiro Yogo, M.. 2005. “Testing for Weak Instruments in Linear IV Regression.” In Stock, J.H., Andrews, D.W.K., (eds.), Identification and Inference for Econometric Models: A Festschrift in Honor of Thomas Rothenberg. Cambridge University Press, Cambridge: 80–108. (Ch. 5).

Taylor, Edward J. 1986. “Differential Migration, Networks, Information and Risk.” In Oe Stark (eds.) Migration Ttheory, Human Capital and Development. Greenwich, Conn.: JAI Press: 147-171.

Todaro, Michael P. 1969. “A Model of Labour Migration and Urban Employment in Less Developed Countries.” American Economic Review 59:138-148

Uhlenberg, Peter. 1973. “Noneconomic Determinants of Nnonmigration: Sociological Consideration for Migration Theory. ” Rural Sociology,38:296-311,

Wang, Guomin. 2007. The Risk Management System of Natural Disaster in Agriculture, Social Science Research :4 (In Chinese)

Wooldridge, J. M. 2006. Introductory Econometrics. A Modern Approach. (Third Edition). Thomson South-Western.

---- 2002，Econometric Analysis of Cross Section and Panel Data. Cambridge，MA：MIT Press.

Wu, Xiaogang, and Donald J. Treiman. 2007. “Inequality and equality under Chinese Socialism: the Hukou System and Intergenerational Occupational Mobility.” American Journal of Sociology, 113 (2): 415–445.

Zhao, Yaohui, 2003, “The Role of Migrant Networks in Labour Migration: The Case of China,” Contemporary Economic Policy, 21(4): 500-511.

----1999. “Leaving the Countryside: Rural to Urban Migration Decisions in China,” American Economic Review 89: 281-286.