Migration With Endogenous Social Networks in China

Jin Zhou
(University of Western Ontario)

May 2015

Abstract

Numerous empirical studies have documented a strong association between social networks and individuals' migration decisions. Few papers formally analyse how social networks affect both migration decisions and labor market outcomes formally. In order to understand how social networks affect individuals' migration decisions and labor outcomes, I develop and estimate a dynamic programming (DP) model for the joint decisions of repeat and return migration, social network investment decisions and labor market outcomes. I use data on internal migration between rural and urban areas in China. The model distinguishes between the two channels: migration cost and search frictions through which social networks may affect migrations. On one hand, social networks may have a direct effect on migration costs. On the other hand, social networks may have an indirect effect on labor outcomes by the impact on job arrival rate. The estimation results show that social networks affect both channels. Individuals with networks have almost twice job arrival rate than those without networks. At the same time, social networks can reduce 10% of migration cost on average. To achieve the same government goal, policy simulations show that the government has to spend more to offset the effect of lower investment in social networks.

KEYWORDS: Internal Migration, Search, Social Networks.

This is a preliminary draft. Please do not cite or distribute without permission of the author.

1 Introduction

A strong association between social networks and migration decisions has been consistently documented in numerous empirical studies. In most of models, migration decisions are based on potential labor market outcomes. Social networks are often viewed as an important factor which may affect these outcomes. However, there are competing findings about how social networks affect individuals' labor market outcomes. For example, social networks may provide access to better jobs (Munshi (2003); Edin, Fredriksson, and Aslund (2003)) or to less desirable ones (Borjas (2000); Chiswick and Miller(2005)). Although some researchers point out that individuals with social networks in destination places are more likely to migrate (i.e. Munshi (2003)), there are few papers to analyse how social networks affect individuals' migration behavior and their labor market outcomes formally.

The existing migration literature suggests two alternative mechanisms through which social networks may affect migration decisions and migrants' labor market outcomes. First, social networks may reduce migration costs (e.g., Carrington, Detragiache, and Vishwanath (1996); Munshi (2003)). Social networks decrease individuals' migration reservation values so ones with networks are more likely to migrate. On the other hand, social networks provide information about labor markets and then increase the probability of getting jobs in the destinations (e.g., Kono (2006); Goel and Lang (2012); Buchinsky, Gotlibovski, and Lifshitz (2012)). It also implies that ones with social networks are more likely to migrate.

Although both of the two mechanisms can explain why ones with networks are more likely to migrate, they may derive different implications about migrants' earnings. Under the first explanation, individuals with social networks have lower migration costs. It means that migrants with networks may have lower earnings compared to those without networks. However, if social networks will reduce the search frictions, it implies that migrants with networks will have higher earnings than those without. Since these two potential mechanisms have different implication about migrants' earnings, understanding which mechanism(s) is the key determinant of migration will help the government to implement migration policies more efficiently.

To understand how social networks affect individuals' migration decisions and labor market outcomes, I construct and structurally estimate a dynamic model of the joint migration choices including repeat and return migration, unemployment, social network investment decisions of men. In the model, social networks may affect migration behaviours and labor market outcomes through two channels: migration costs and job arrival rates in urban areas. Also, social networks can be determined endogenously by social network investment.

Compared with most research considering social networks as exogenously given or choosing to use natural experiment or quasi-natural experiment to deal with the endogenous problem of social networks, this paper endogenizes individuals' social network investment[1]. Individuals will make choices of investing their social networks by comparing the monetary loss from network investment and the benefit from increasing the probability of having social networks. Modelling social networks with endogenous investment is help to understand how individuals change their investment decision and ultimate their migration and labor market outcomes. It is also essential to endogenize social networks when evaluating the government policies.

The current internal rural-urban migration in China provides a nice background to examine the role of social networks in a labor market with frictions. As shown in Hare and Zhao (2000), Meng (2000) and Zhao (2003) social networks are strongly correlated with rural-urban migration in China. Zhang and Zhao (2011) find social networks affect migrants' subsequent labor market outcomes. More than 140 million rural people migrated to urban cities by 2008, this fraction is 32% of the total number of rural laborers[2]. However, the government aims to increase the urbanization rate to 60% by 2020, which means that an additional 100 million rural people need to migrate to urban cities. The model developed in this paper can be used to evaluate several types of policies that might be used to answer this desired outcome.

This paper is the first study to examine the role of social networks through both channels (i.e. migration cost and search frictions) in order to explain the mechanisms underlying the observed correlation between migration and networks. It is also the first study to allow individuals endogenously to invest their social networks. At the same time, it is also the first study to consider rural-urban migration in China in a dynamic setting with frictions.

Besides the impact of social networks, the model in this paper contains a number of mechanisms through which individuals' migration decisions are affected. First, this model allows individuals to accumulate human capital within a search framework. Individuals' earnings reflect their characteristics that may be either observed (e.g., education) or unobserved (e.g., ability) to the econometrician. Earnings also reflect their location-specific human capital accumulation (i.e. urban and rural), which I model as a learning by doing process in which human capital is accumulated based on location specific work experience.

Second, individuals' earnings are also affected by frictions in the urban labor mar-

[1] Goel and Lang (2012) consider to examine how social networks affect labor outcomes but they do not allow agents to have their own decisions to invest social networks. They use difference-in-difference strategy to avoid endogenous issues.

[2] rural laborers mean the adults whose ages are between 16 and 60.

ket. They do not automatically have a job if they migrate, instead they need to search for one. Depending on the outcome of the search process, individuals may choose to stay in urban areas or return to rural areas. Allowing for the possibility of return and repeat migration leads to a dynamic model.[3] Failure to account for the dynamics of job search, and return and repeat migration may lead to biased estimates and invalid inference.

In this paper, I estimate the model by maximizing the likelihood function. From the estimation results of exogenous and endogenous social network models, I find actually networks affect on both migration costs and job arrival rates. Individuals with networks have almost twice job arrival rates than those without. Social networks also can reduce 20% of migration costs. To analyse the role of these two channels, decomposition studies show that migration decisions are more influenced by whether social networks reduce search frictions. In the endogenous model, if social networks do not affect both channels, only 17% of rural people will migrate. If allowing social networks only affect migration costs, addtional 2% of people will migrate. If allowing social networks increase job arrival rate, additional 10% of rural people will migrate. The exogenous model has the similar findings.

The decomposition results also show how individuals respond to the impact of social networks by network investment. When social networks affect both channels, the network investment rate is 26%. If social networks only lower migration costs, the investment rate decrease to 6%. When social networks' impact is only through job arrival rate, individuals' network investment fraction is 23%.

Since the government tries to increase the urbanization rate to 60% by 2020, I try three different policies to achieve the same goal (i.e. 60% urbanization rate). The simulation results show that lump-sum transfer will spend less for the government. When I compare the two different models, I find the government has to spend more if individuals can endogenously invest their social networks. The reason is that individuals do not invest more for their networks and the government has to spend more to offset the effect of lower network investment.

The rest of this paper is organized as follows. Section 2 provides a review of relevant literature. Section 3 presents the data. In Section 4, the model is described and identification condition and estimation methods are given. Estimation results and counter-factual studies are shown in Section 5. Section 6 concludes.

[3]More than 45% rural migrants had the experience of return and repeat migration.

2 Literature Review

There are two findings consistently found in the migration literature. The first finding is that individuals with social networks are more likely to migrate for rural-urban migration in China (i.e Hare and Zhao (2000); Meng (2000) and Zhao (2003)). The second finding is that social networks affect migrants' subsequent labor market outcomes. For example, Zhang and Zhao (2011) examined the correlation between social-family networks and self-employment. They find social-family networks increase migrants' employment probability in urban cities. However, little is known about the role of social networks in China in determining subsequent labor market outcomes like earnings, job arrival rates, spells of job search.

In current migration studies, most of them consider that heterogeneous migration costs explain why some individuals migrate and some do not. However, there are few formal studies to examine whether social networks affect migration costs. Carrington, Detragiache, and Vishwanath (1996) build a dynamic model to analyse the phenomena that more black people migrated from South to North of US when they faced a smaller wage gap in Great Migration period. They claim that although the income gap was larger before 1930s, black people did not have social networks in North part of US and migration costs were large. They show that social networks can influence individuals' migration decisions since they might have lower migration costs if they have social networks in the destination place. However, they do not quantitatively examine how social networks affect migration costs, and assume that each individual has homogeneous social networks. They do not distinguish search frictions from migration costs.

Besides the friction of existing migration costs, search frictions in the destination labor markets will also affect individuals' migration decisions. Gemici (2011) compares migration behaviours between married couples and singles in a dynamic model of household migration with bargaining between family members. In her model, she allows that there exists uncertainty in labor market and individuals' migration decisions are influenced by search frictions. She finds that migration of married couples occurs much more in response to the earning of men than to the earning of women, as women have lower wage offers, and a lower arrival rate of offers. Buchinsky, Gotlibovski, and Lifshitz (2012) examine the effect of a few alternative national migration policies on the regional location choices and labor market outcomes of migrant workers. In their paper, they estimate a dynamic programming discrete choice model with incorporating stochastic job offers and job terminations. But these studies do not consider how social networks affect the search frictions.

My study is also related to several papers that have analysed that individuals with networks are more likely to have a better labor market outcomes. Munshi (2003) follows Carrington, Detragiache, and Vishwanath (1996)'s idea to examine how social networks affect Mexican migrants to the US. Since the size of social networks is endogenously determined, he uses last period rainfall as the instrument and finds that individuals are more likely to be employed and to hold a higher paying non-agricultural job when the size of network is exogenously larger. However, this study assumes that the probability of being employed in the destination can be independent with the individuals' duration at the destination which rules out individuals can endogenously invest their networks and try to reduce search frictions. The theory paper, Kono (2006) shows that workers with social networks have fewer information deficiencies because they can use referral channels to find a job. Therefore, individuals with social networks will have higher wages than those without. However, from the empirical analysis results, it is not clear whether social networks could have a positive effect on individuals' wages. For example, social networks may provide access to better jobs (Munshi (2003); Edin, Fredriksson, and Aslund (2003)) or to less desirable ones (Borjas (2000); Chiswick and Miller(2005)).

Goel and Lang (2012) examine how social networks affect immigrants' labor market outcomes. In their model, the mechanism is that social networks can increase the probability to get a job offer. To avoid the endogenous problem of the size of social networks, they employ the difference-in-difference approach.

3 The Background, Data and Preliminary Examination

3.1 The Background of Rural Urban Migration in China

Since 1958, the Chinese central government began to restrict the mobility of the population. From 1958 to 1983, only the rural people who have the job offer in urban cities or recruitment letter from universities could migrate from rural to urban areas. The central government began to allow rural urban migration but rural agents have to provide food for themselves between 1984 and 1988[4]. Although the government released the restriction of migration, it was still hard for rural individuals to migrate since it was not easy to have enough food stamps to support themselves. This migration policy had been prohibited between 1989 and 1991. After 1992, the government began to encourage rural urban migration and since 2000, the government started to reform household registration system to encourage more rural individuals to migrate. For example, in 2007, 12 provinces in China had cancelled the rural household registration, which means that rural individuals can have the same household registration

[4] At that time, it was planned economy in China. The amount of food for each individual is planned by the government. People need to use food stamp to exchange food.

as urban households in these provinces.[5]

The government policy has a significant effect on agents' migration decisions. Table 1 gives the inter-provincial migration in China from 1990 to 2005. There are 9.2 million people who migrate inter province between 1990 and 1995 and this number increases into 32 million between 1995 and 2000 and 38 million between 2000 and 2005.

3.2 Data

This study uses the first three waves (2007-2009) of the China Household Income Project (CHIP) panel survey. This database is planned to be a five-year panel survey in China with the goal of studying issues such as the effect of rural-urban migration on income mobility and poverty alleviation, the state of education and health of children in migrating families.

Three representative samples of households were surveyed, including a sample of 8000 rural households, a sample of 8000 rural migrant households , and a sample of 5000 urban households in 9 provinces. The 9 provinces in the survey cover the most important provinces of the migration origin and destination in China. Figure 2 gives a map of the 9 provinces. Table 1 shows that from 2000 to 2005 more than 68% of migrants moved into those 9 provinces while 52% of migrants moved out of those 9 provinces (NBS 2002, 2007).

In the analysis, I use the CHIP rural male samples, which contain information on work experience, job search durations, work locations, earnings, the presence of social network and social network investment decisions. Using this data, I can construct the location choices, job search duration, and work status for the individuals who are between 16 and 60 years old for the three year period.

I do not use the migration sample in this paper. First, the response rate in the migration data is quite low. Attrition rate is above 70% to build three years panel data. Second, I cannot follow the history of migrants work experience using the migration samples. For example, migrants who return to their home towns are not surveyed. In contrast, using the rural sample data, I can build the full history of the work experience for three years no matter where the agents are located.

The total number of individuals for the 8000 households is 33,396 with 16,583 males. After restricting age to be between 16 in 2007 and 60 in 2009, the sample size shrinks to 11,385. In the data, there are 1030 observations missing the information on fertility

[5] These 12 provinces are Hebei, Liaoning, Jiangsu Zhejiang, Fujian, Shandong, Hubei, Hunan, Guangxi, Chongqing, Sichuan and Shanxi.

decisions or marital status. There are 1099 observations with missing work experience information during 2007 to 2009. The sample used for estimation includes 9,256 males in 6400 households. The panel is balanced except the social network investment choices. Only the first two years' data contain the information about social network investment.

Social networks are defined as the presence of friends or relatives who living in urban areas, contacted by households. The social network investment are the monetary values of gifts to their friends or relatives. In the survey, people will answer whether to give gifts to your friends or relatives and also give the monetary values of gifts. In the data, the gifts could be given to the friends or relatives who living in rural areas. At the same time, individuals may build social networks through other channels (i.e. call each other, take care of friends' children or older family members). These two possibles bring the two side measurement errors of the variable social network investment. Estimation section provides details to show how to deal with measurement error problem.

Table 2 displays selective descriptive statistics of the sample used in estimation. The average size of social networks is 6.8 persons. The size of social networks is self-reported, which are more likely to be the number like 5, 10, 15. Figure 3 gives the density of self-reported network size. To avoid this problem, I categorized the self-reported size into 5 categories (i.e. 0, 1-6, 7-11, 12-21, and > 21). More than 30% of individuals who live in households with no social networks in urban cities. In the data, households report the monetary value of gifts they sent to their friends or relatives. More than 60% of individuals try to invest their social networks in 2007 and around 77% invest in 2008.

All individuals finished their formal education. The average education year is 8.3 years. Since 1989, mandatory education in China is 9 years which is equivalent to complete middle school (or finish primary school). More than 80% of individuals have finished middle school or lower. Monthly earnings in urban areas are almost five times the earnings in rural areas.

The definition of migration is whether the urban residence location is out of his (her) rural *hukou* (household registration) county. Table 3 displays the descriptive statistics between migrants and non-migrants. Migrants have better education levels in general than non-migrants. The education levels are higher for the agents with networks than those without networks among both migrants and non-migrants groups. The average age of migrants is 31, which means that migrants are much younger than non-migrants. Non-married ones are more likely to migrate. Migrants with social networks have higher earnings than those without social networks. At the same time, migrants with networks have a smaller variance of the earnings. Non-migrants' earnings do not have significant differences between the two group of individuals. Migrants with networks have a slightly longer job search duration than those without.

3.3 Preliminary Examination

Before introducing the structural model, in this section, I examine several correlations which are the important mechanism of this paper. Firstly, I document that there exists the strongly correlation between social networks and migration choices and subsequence labor outcomes.

Table 6 shows two examinations: the column 1 gives the correlation between migration choices and the size of social networks and the correlation between networks and earnings are shown in column 2. From column 1, conditional on education levels, marital status, number of kids and age, individuals with a larger size of social networks are more likely to migrate. Column 2 shows that employed migrants with social networks have higher urban earnings than those without. From table 6, it is clear that social networks are significantly correlated with migration decisions. Also, migrants with networks have higher earnings even after controlling for education, marital status and the number of children.

Since social networks are a key factor to impact individuals' migration choices, individuals can invest their social networks to increase the size of social networks. Next, I examine whether there exists a strong association between social network investments and the size of social networks.

Table 4 gives the correlation between the mobility of social networks and social network investment. Table 4 shows social network transition matrix conditional on social network investment choices. The columns are the size of social networks in 2008 and rows are the size of social networks in 2009. From Table 4, when comparing the numbers between part A and part B, if individuals live in the household that invest in their social networks in 2008, the fraction of individuals keeps the same level or has larger size of networks is bigger than individuals whose households did not invest social networks in 2008. The Part C in Table 4 shows that individuals in households with investments are more likely to have a larger size of social networks in the next period. Most of under-triangle values are positive numbers.

Table 5 gives the examination results of the correlation between social networks and network investment by the dynamic ordered probit model. The coefficient of social network investment choices is positively significant which is consistent with the findings in Table 4. That means if the individual invest his (her) social networks at the period $t-1$, the probability of getting the larger size of social networks at period t is increasing.

The central government migration policy affects individuals' location choices. Individuals in different cohorts show different migration patterns. Figure 5 shows that the fraction of individuals who migrate to urban cities from 2007 to 2009 is increasing linearly across different cohorts. Figure 6 examines average ages of individuals at the first time migration across different cohorts. It shows a clear pattern that average ages of first time migration are decreasing linearly with cohorts. Both figure 5 and 6 show that individuals in different cohorts have different migration patterns. In the model, I introduce the cohort effect in the migration cost function incorporating the government policies by which different cohorts faced.

Next, since the government policies are changing over time, the natural question whether the year effect is also an important factor. The survival analysis is examined to see whether both cohort and the year effect affect average ages of individuals' first time migration. Table 7 gives the estimates with the assumption of the loglogistic distribution. The coefficient of education shows that individuals with higher level of education will migrate earlier. The year dummies are the time when the central government made a big change about migration polices. The year dummies (1984-1991) and (1992-2000) are not significant at 5% levels. Only the year dummy (2001-2009) is significant. Comparing with cohort effects, year effects do not have a strong impact on individuals' migration choices.

4 Model

I model individuals' migration decisions in a finite-horizon framework. To consider the uncertainty regarding wage offers, the migration decision problem is incorporated into a dynamic discrete time search framework. The decision period in this paper is one month length. In each period, men receive flow utility associated with their current locations and incur a migration cost if rural individuals decide to migrate to urban cities.

The timing of decision process is important and emphasizes the effect of expectations on migration decisions. Individuals know their flow utility associated with the current location and employment states. If they move from rural to urban areas, they pay all migration costs which is known by individuals in the current period. If they return from urban areas, they also need to pay return migration costs. Migration costs are allowed to be different from return migration costs.

Individuals' flow utility comes from their earnings, unemployment benefits, physic values of living in home towns, and an additively-separable choice specific shock. Migration costs are specified as a function of marital status, the number of kids, the presence of social networks (i.e. friends or relatives) in urban areas and individuals' birth cohorts. The agent lifetime utility is given by current utility flow and the discounted stream of expectation future utilities. Uncertainty in this paper comes from search frictions in urban labor market, the transition of earnings and idiosyncratic shock to utility.

4.1 Objective

This paper tries to answer how social networks affect individuals' migration behaviours through two frictions: migration costs and search frictions in urban markets. Although in this paper social networks do not affect individuals' earnings directly, social networks may affect earnings by changing job arrival rates in urban cities. Also, social networks may lower migration costs which is commonly documented by migration literature. Most migration literature state social networks can reduce migration costs but ignore social networks may also decrease search frictions in labor markets.

The following toy example analyses why it is necessary to consider the impact of social networks through two channels (i.e. migration costs and search frictions). Figure 7(a) gives the urban earnings' population for rural individuals and the black line is the migration cost. Only the individuals whose urban earnings are larger than the migration costs want to migrate. If social networks only lower migration costs, for example, in Figure 7(b), migrants with social networks have lower earnings compared with those without social networks. However, if social networks only reduce search frictions, Figure 7(c,d) shows that migrants with social networks have higher earnings compared with ones without social networks.

These two mechanisms give the opposite predictions between social networks and migrants' earnings. The opposite predictions also consistent with empirical findings: Munshi (2003) and Edin, Fredriksson, and Aslund (2003) find individuals with social networks have better labor outcomes; Borjas (2000) and Chiswick, Lee, and Miller (2005) support that migrants with networks have less desirable works.

This paper tries to consider both mechanisms (migration costs and search frictions) and clearly analyses the role of social networks through these two different channels, and examine how social networks affect migration and subsequent labor market outcomes.

Since the current Chinese government tries to encourage internal rural-urban migration, I can use the estimation results to do counterfactual analysis. For example, I can compare different polices in terms of the government budgets to migrate a given amount of rural individuals. This paper also examines whether it is possible to achieve the target of urbanization rate (i.e. 60% in 2020) without the government intervention.

4.2 Basic Structure

4.2.1 Earnings

Individuals receive their earnings which are functions of education and work experience in rural and urban areas and location idiosyncratic shocks on earnings. The earnings of individuals i in location $j \in \{u, r\}$ (u:urban, r:rural) at time t are described as

$$e_{it}^{j} = \beta_1^j S_i + \beta_2^j \text{exp}_{it}^{r} + \beta_3^j \text{exp}_{it}^{u} + \beta_4^j \text{exp}_{it}^{r\,2} + \beta_5^j \text{exp}_{it}^{u\,2} + \beta_6^j \theta_i + \varepsilon_{it}^{j} \tag{1}$$

where S_i is education years. In this paper, individuals accumulate their human capitals through learning by doing by location-specific work experience (i.e. rural exp^r, urban exp^u). θ_i is the type of individuals which is unobserved by econometricians. θ_i reflects the unobserved factor which may affect individuals' productivity (i.e. ability). Here work experience in rural and urban areas are dependent on the history of endogenous decision $\{d_{ik}\}_{k=1}^{t-1}$, which includes location, employment, and network investment choices. Shock terms ε_{it}^{j} are i.i.d. across individuals, locations, and time and they are normally distributed with mean zero and variance σ_j^2. Individuals know the current period transient component. However, they do not know values of future transient components but they know distributions.

The effect of social networks on earnings, especially in urban areas, is not reflected directly from earning equations. However, social networks may affect earnings indirectly. For example, individuals with social networks may have higher job arrival rates which will increase reservation values of taking urban job offers. If so, individuals with social networks may have higher accepted earnings; on the other hand, social networks may reduce migration costs. Hence individuals with social networks are more likely to migrate because of lower migration costs. It implies that individuals with social networks may have lower reservation values of taking urban job offers compared with those with fewer social networks. From the discussion above, the direction of how social networks affect earnings is not clear since social networks may play different roles (i.e. lower migration costs and increase job arrival rates) at the same time.

4.2.2 Migration and Return Migration Costs

If individuals migrate from rural to urban areas, they have to pay migration costs. One of key assumptions is that social networks in urban cities may affect their migration costs. In Carrington, Detragiache, and Vishwanath (1996), they build a dynamic macro model to examine the role of social networks on migration decisions. They also assume social networks reduce migration costs.

Migration costs M_{it} depend on the current period's social network status, marital status, the number of children, birth cohort and individuals' type. Since individuals have more information about their own home towns, I assume asymmetric migration costs: migration costs may not be equal to return migration costs. Migration and return migration costs in this paper are specified as the equation 2-3:

$$M_{it} = \beta_1^m 1_{sn_{i,t}} + \beta_2^m mar_{it} + \beta_3^m child_{it} + \beta_4^m cohort_i + \beta_5^m \theta_i + \varepsilon_{it}^m \tag{2}$$

$$RM_{it} = \beta_1^{rm} 1_{sn_{i,t}} + \beta_2^{rm} mar_{it} + \beta_3^{rm} child_{it} + \beta_4^{rm} cohort_i + \beta_5^{rm} \theta_i \tag{3}$$

where $1_{sn_{i,t}}$ is the indicator of the presence of social networks in urban areas at period t. In this paper, the variable of social networks shows the presence of contacting relatives or friends who living in urban areas within the year. Migration costs may be also relative to their birth cohorts, since they may perform differently to assimilate the new environment and face different migration policies.

4.2.3 Job Arrival and Separation rate

This paper assumes that there exists search friction in the urban labor market. If rural people migrate to urban areas, they have to search jobs from unemployment state. Social networks may help individuals to reduce search frictions in urban areas. So the job arrival rate λ_{it} in urban areas is parametrized as:

$$\lambda_{it} = \frac{\exp\{\beta_1^l 1_{sn_{i,t}} + \beta_2^l S_i + \beta_3^l \theta_i\}}{1 + exp\{\beta_1^l 1_{sn_{i,t}} + \beta_2^l S_i + \beta_3^l \theta_i\}} \tag{4}$$

In the urban labor market, a significant fraction of migrants do not have long term contract for their jobs and firms have low costs to fire rural workers. To model exogenous job separation, job destruction rate is parametrised as:

$$\delta_{it} = \frac{\exp\{\beta_1^\delta S_i + \beta_2^\delta \theta_i\}}{1 + exp\{\beta_1^\delta S_i + \beta_2^\delta \theta_i\}} \tag{5}$$

4.2.4 Social Network and Investment Decision

Social Network Formation:

In the model, social networks are determined endogenously. Individuals can pay social network investment τ_{it}^* to keep connections with their friends (i.e. they may give gifts to their friends or contact with their friends by phone, mailing). Social networks are formed by the following dynamic ordered probit model:

$$sn_{it}^* = \beta_1^s inv_{i,t-1} + \beta_2^s mar_{it} + \beta_3^s child_{it} + \beta_4^s 1_{sn_{i,t-1}} + \beta_5^s \theta_i + \varepsilon_{it}^s \tag{6}$$

where $inv_{i,t-1}$ is individual's endogenous investment decision at period $t-1$. $inv_{i,t-1} = 1$ means that she decides to invest her social networks at period t, otherwise $inv_{i,t-1} =$

0. She can increase the probability of having social networks by the investment choice (i.e. giving gifts). $1_{sn_{i,t-1}}$ is the indicator of social networks status in period $t-1$. The shock ε_{it}^{s} is i.i.d. across individuals and time. Individuals cannot observe shock terms ε_{it}^{s} but they know the distribution of shocks. Hence, they can increase the probability of getting social networks by network investment $inv_{i,t-1}$.

Investment Decision:

Social network investment decision here may increase the probability of getting social networks. In this model, investment decision is discrete choice that depends on the trade-off between the gain from increasing the probability of having larger size of networks and the cost of investing.

4.2.5 The Value of Unemployment in Urban Cities

Values of being unemployed in urban areas are considered as a function of age, marital status and the number of children. Elder people may get hard to assimilate to new environment so they may have different evaluations of being unemployed in urban areas. Marital status and the number of children will reflect net values between utility of living with family and costs of living with family.

$$\xi_{it} = \beta_1^{\xi} age_{it} + \beta_2^{\xi} age_{it}^2 + \beta_3^{\xi} mar_{it} + \beta_4^{\xi} child_{it} + \beta_5^{\xi} \theta_i + \varepsilon_{it}^{\xi} \tag{7}$$

where ξ_{it} is the value of unemployment in urban areas for individual i at period t which is a function of age, marital status, the number of children and the agent i's unobserved type.

4.2.6 Marriage and Fertility Transition Process

In this paper, I model individuals annually marriage and fertility transition process by exogenous processes. Individual's marital transition process is modelled by the continuous duration model with loglogistic distribution. The survival function is:

$$Sur_i(t) = (1 + (e^{-(\beta_1^{ma} S_i)} t)^{1/\gamma})^{-1} \tag{8}$$

Here, $Sur_i(t)$ is the probability of being single at age t, S_i is education year for individual i, and γ is the parameter of logogistic distribution. Then, conditional probability of getting married at period t:

$$\Pr(mar_{it} = 1 | mar_{i,t-1} = 0) = \frac{Sur_i(t) - Sur_i(t+1)}{Sur_i(t)}$$

The fertility decision is made based on the following equation:

$$F_{it} = \begin{cases} 1 & \text{if} \quad \beta_1^f age_{it} + \beta_2^f age_{it}^2 + \beta_3^f child_{it} + \beta_4^f child_{it}^2 + \beta_5^f S_i + \beta_6^f mar_{it} + \varepsilon_{it}^f > 0 \\ 0 & \text{else} \end{cases} \tag{9}$$

This equation shows that fertility decision is a function of age, the number of children, marital status and education.

4.2.7 Psychic Value

People may value their home towns differently and older persons may have deeper homesick. At the same time, the utility from marriage and children may also affect psychic value of living in rural areas. In this paper, the amenity of home location is modelled by the following equation:

$$\phi_{it} = \beta_1^\phi age_{it} + \beta_2^\phi age_{it}^2 + \beta_3^\phi mar_{it} + \beta_4^\phi child_{it} + \beta_5^\phi \theta_i \tag{10}$$

4.2.8 State Space

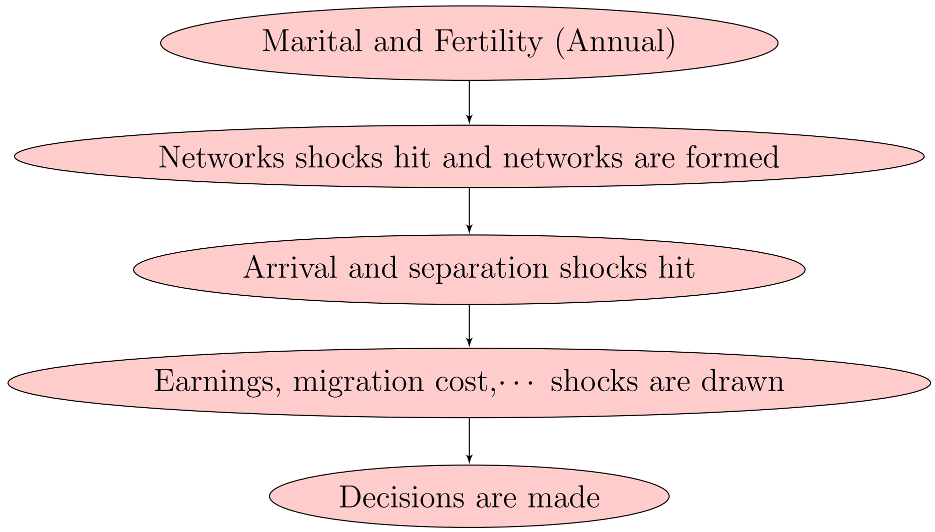
The vector of state variables for individual i at time t is denoted as H_{it}. State variables for a given time t includes age, education year, marital status, number of children, accumulated work experience in rural and urban areas, the presence of social networks and social network investment at period $t-1$. Control variables include individuals' decisions (i.e. migration, employment in urban areas, employment in rural areas, unemployment in urban areas, return migration, and social network investment decisions).

I assume that the transition of state variables is Markov chain, and then $H_{i,t+1}$ depends on $H_{i,t}$ and D_{it} only; no additional information is gained from the information $H_{i,t-1}$ and the transition probability of state variables $H_{i,t}$ is denoted as $\Pr(H_{i,t+1}|H_{i,t}, D_{i,t})$.

The transition of social networks is given by the dynamic probit model. Work experiences in rural and urban areas is determined by the action history $D_{it} = \{\{d_{it}^k\}_{t=1}^T, inv_{it}\}$, where $\{d_{it}^k\}_{t=1}^T$. $k \in \{1, \cdots, 6\}$ (1: migrate, 2: employed in urban, 3: employed in rural, 4: unemployed in urban, 5: return migrate, and 6: social network investment).

4.3 Timing of the decisions

Individuals' choices are made sequentially and also based on their current locations. Here, before describing the value functions, I specify the timing of individuals' decisions.



The Timing of Decisions

1. Individuals draw the marital and fertility shocks and then they make marital and fertility decisions annually.

2. At the beginning of period t, the shock for social networks realised monthly and then individuals observe their social networks at period t.

3. Job arrival and separation shocks hit.

4. Earnings, migration cost, and unemployment benefits shocks are drawn.

5. Following all of these shocks, location,employment and network investment choices are made

4.3.1 Timing based on location

In this paper, individuals make decisions based on their locations. Let $W_{it}^{j}(inv_{it})$ be the value of state $j \in \{e, r, n\}$ (e: employment in urban areas, r: rural, n: unemployment in urban areas). inv_{it} is the indicator of network investment. Denote $V_{it}^{j} = \max\{W_{it}^{j}(1), W_{it}^{j}(0)\}$. I will describe the decision process separately:

If the individual is in a rural area, her earning e_{it}^{r} is drawn from the distribution $G(e^r)$. She knows her migration cost M_{it} and the value of unemployment in urban area. Hence, the migration decision is made based on the following equation:

$$D_{it}^{mig} = \begin{cases} 1 & \text{if} \quad \max\{W^n(1)_{it} - M_{it} - \nu, W^n(0)_{it} - M_{it}\} \geq \max\{W^r(1)_{it} - \nu, W_{it}^r(0)\} \\ 0 & \text{else} \end{cases}$$

where ν is the cost of network investment.

Network investment choice is made:

$$Inv_{it} = \begin{cases} 1 & \text{if} \quad \max\{W^n(1)_{it} - M_{it} - \nu, W^r(1)_{it} - \nu\} \geq \max\{W^n(0)_{it} - M_{it}, W^r_{it}(0)\} \\ 0 & \text{else} \end{cases}$$

If the individual is in an urban area, which means she has already migrated, her choices are made based on the following conditions:

1. If she just arrived in urban areas, she has to be in the unemployment state for one period (month) and gets unemployment value ξ_{it}.
2. If she has already stayed for one or more periods in urban areas, she can get a job offer with probability λ_{it}, which is affected by the presence of social networks $1_{sn_{i,t}}$.
   (a) If she gets a job offer, she will make a decision between three options, i.e. unemployment in urban areas, taking the job offer, or return migration $(\max\{V^n_{it}, V^e_{it}, V^r_{it} - RM_{it}\})$. Social network investment choice is made at the same time.
   (b) If she does not get a job offer, (with probability $1 - \lambda_{it}$), she will select between two options i.e. unemployment in urban area or return migration $(\max\{V^n_{it}, V^r_{it} - RM_{it}\})$. Social network investment choice is made at the same time.
3. If the individual works in urban city, the exogenous job separation happens with probability δ_{it}. She will choose to search job in urban areas or to return migrate and network investment choice.[6] If the separation shock does not hit her, she will choose to keep the job, quit the job with unemployment in urban city or quit the job and return to their home locations and network investment choice.

[6] I do not consider on-the-job search. The fraction of on-the-job search is quite low which is less than 2% for who quit, switch jobs.

4.4 Value Function

The Bellman equations for each of the three states are

$$
\begin{aligned}
W^n_{it}(inv_{it}; H_{it}) &= \xi_{it} - \nu inv_{it} + \frac{\lambda_{it}}{1+\rho} E(\max\{V^n_{it+1}(H_{it+1}), V^e_{it+1}(H_{it+1}), V^r_{it+1}(H_{it+1})\}|H_{it}) \\
&\quad + \frac{1-\lambda_{it}}{1+\rho} E(\max\{V^n_{it+1}(H_{it+1}), V^r_{it+1}(H_{it+1})\}|H_{it}) \\
W^e_{it}(inv_{it}; H_{it}) &= e^u_{it} - \nu inv_{it} + \frac{\delta_{it}}{1+\rho} E(\max\{V^n_{it+1}(H_{it+1}), V^r_{it+1}(H_{it+1})\}|H_{it}) + \frac{1-\delta_{it}}{1+\rho} \\
&\quad E(\max\{V^e_{it+1}(H_{it+1}), V^r_{it+1}(H_{it+1}), V^n_{it+1}(H_{it+1})\}|H_{it}) \\
W^r_{it}(inv_{it}; H_{it}) &= e^r_{it} - \nu inv_{it} + \phi_{it} + \frac{1}{1+\rho} E(\max\{V^r_{it+1}(H_{it+1}), V^n_{it+1}(H_{it+1}) - \\
&\quad M_{it+1}(H_{it+1})\}|H_{it})
\end{aligned}
$$

where, ξ_{it} is relative expenditure and benefits in the state of unemployment in urban areas. e^u_{it} are earnings in urban areas and e^r_{it} are rural earnings. M_{it} is migration cost, τ_{it} is social network investment and ϕ_{it} is psychic value.

The state variables H_{it} include education S_i, work experience in rural areas $\exp^r_{it}$, work experience in urban areas $\exp^u_{it}$, marriage indicator mar_{it}, children number $child_{it}$, the presence of social network $1_{sn_{it}}$ and social network investment $inv_{i,t-1}$ at period $t-1$.

4.5 Identification

The model in this paper is a partial equilibrium model. I assume that the offered earnings' distributions are log normal. Based on the normality assumption, the variance term of earnings' distributions can be identified since we observe the accepted earnings. The distribution of unemployment value shocks in urban cities is assumed to be normality. The variance term of unemployment value shock σ^ξ can be identified from the probability of returned individuals given the variances of earnings' distributions (i.e. σ^r, σ^u).

Job arrival rates λ can be identified by the variation of unemployment job search duration in urban areas. Since the fraction of employed migrants who switch into unemployment state can be observed, job separation rate δ can be identified once the variance of σ^ξ, σ^r and σ^u are known.

Psychic value ϕ_{it} of living in rural areas and migration costs M_{it} can be separately identified. The reason is that agents only pay migration costs when they actually do migrate while individuals have the utility of psychic value every period. This setting is naturally to help to separate identify these two terms. Also, there exists a exclusive variable: birth cohort in migration cost function which also helps to identify migration costs from the flow of psychic value. The discount factor ρ is not identified and I set $\rho = 0.0025$ which is target to 3% (annul interest rate).

The unobserved type can be identified from the consistent behaviour observed from panel data. For example, through earning equations, conditional on the same observables if two agents have persistent large earning gaps, it will provide the identification of unobserved type. The loadings of unobserved type can be identified by the correlated relationship of residuals after controlling for observables.

4.6 Objective Function and Solution Method

Individuals maximize their present discounted value of lifetime utility from the year since which they finished their education to a terminal age, $t = T$. Each individual chooses a choice d_t^k, $k \in \{1, \cdots, K\}$. Denote the utility associate with the choice as u_t^k. Then individuals choose the choice to solve the following maximized objective function $V_{it}(H_{it})$:

$$V_{it}(H_{it}) = \max_{D_{it}^k} u_{it}^k(H_{it}) + \frac{1}{1+\rho} E(V_{it+1}(H_{it+1})|H_{it}) \tag{11}$$

The expectations operator E in equation(11) is taken with respect to the joint distribution of stochastic shocks $\varepsilon_{t+1}^{\xi}, \varepsilon_{t+1}^{u}, \varepsilon_{t+1}^{r}, \varepsilon_{t+1}^{m}$ and ε_{t+1}^{s} ,the probability of receiving a job offer, job separation, getting married, having a child and the probability of investing social networks.

The solution of the model in general is not analytic. Given the finite horizon, the model is solved numerically through backward recursion on a Bellman equation. The difficulty with this procedure is due to high dimensionality. Since the decision period is monthly, there is computation burden in terms of computation time and memory. To reduce the computation burden, I adopt an approximation method developed in Keane and Wolpin (1994). Specifically, the Emax functions are assumed to be a general function of state space elements. In current version, I use the polynomial functions. For example, in each step, at each t, I calculate the Emax functions for a subset of state space and estimate a regression function as a polynomial in those state space elements. Then using the predicted values from the regression to approximate the alternative-specific value functions given by equation (11).

4.7 Estimation

4.7.1 Likelihood

The model is estimated by the method of maximizing the likelihood function. For each individual, the data consist of the set of choices and outcomes:

- Choices: location, employment, and social network investment (i.e. $\{D_{it}^k : k = 1, \cdots, K\}$)
- Outcomes: earnings, presence of social networks and mobility

for all $t \in [t_{2007}, t_{2009}]$, where t_{2007} is individuals' age at the beginning of the year 2007 and t_{2009} is individuals' age at the end of the year 2009.

Let $c(t)$ denote the combination between choices (i.e. migration, employment, return migration and network investment) and outcomes at each period t and let $\bar{H}_{i,2007}$ denote as the state variables at the beginning of the year 2007. In this paper I assume shock terms are i.i.d across individuals and time, so the probability of any sequence of choices and outcomes can be written as follows:

$$\Pr(c(t_{ini}), \cdots, c(t_{2009})|\bar{H}_{it}) = \Pi_{t=t_{2007}}^{t_{2009}} \int \Pr(c(t)|H_{it}, d_{it}) \Pr(\bar{H}_{2007}|\theta_i) f(\theta_i) \tag{12}$$

where θ_i is the type of individual i.

To calculate the likelihood for each individual i, the whole history of decisions should be known. However for some individuals, the decisions are only observed since the beginning of the year 2007. It is necessary to know the whole history decisions to construct the likelihood for those individuals. To solve the missing history problem, this paper uses simulation method to infer individuals' decision histories. Given the whole histories of decisions are known, the conditional probability of $\Pr(c(t)|H_{it}, d(t))$ are calculated. Likelihood function can be constructed based on the equation 12.

4.7.2 Initial Condition Problem

As I discussed above, to calculate the likelihood for each individual, I need the state variables at the year 2007. The data provide the whole marriage history and fertility history for each individual. However, I miss the information about working experience in rural and urban areas for some individuals. To solve the missing history problem, I use the simulation method. The basic idea is that I simulate the transient shocks and use the value function to simulate individuals' sequential decisions until the time goes to the year of 2007. Then I repeat this simulation procedure many times. After that I calculate the probability of initial conditions for the individuals who miss work experience information. The specific procedure is :

1. For each individual, I draw randomly from the distributions of shocks and then calculate the value function V_{it}
2. Based on the value function, the decision D_{it} is determined and then update the state variables based on the decisions.
3. Given the updated state space, new transient shocks are drawn and I simulate this procedure until t gets to the age at the beginning of 2007.
4. Repeat the procedure from step 1 to step 3 for R times[7], and then I can calculate the probability of each potential state at the beginning of year 2007.

4.7.3 Estimation Procedure

The whole estimation algorithm is developed to incorporate the problem of initial condition problem and to consider the exogenous stochastic process (i.e. marital and fertility process). The whole procedure is described as following:

1. Estimate the exogenous marital and fertility stochastic process and get the parameters Θ_1
2. Given Θ_1, and the initial guess for other parameters Θ_2, the approximation method (i.e. Keane and Wolpin (1994)) is applied to calculate the value functions
3. For the individuals who miss work experience information, I draw the shock terms, and simulate their choices for the missing periods.
4. Repeat the step 3 for R times and calculate the probability of potential initial state
5. Calculate the likelihood and update the parameters Θ_2
6. Repeat from Step 2 to Step 5 until parameters Θ_2 converge.

[7] In my estimation, I simulate 500 times for each individual who miss work experience information.

5 Estimation Results

I estimate two versions of models: the model assumes social networks are exogenously given by dynamic probit process; the other model considers individuals can endogenously increase their social networks through investing their networks. The current estimates do not consider heterogeneous unobserved types.

The estimated parameters are reported in Table 8 (exogenous version) and 9 (endogenous version). Married individuals have higher unemployment values in urban cities and the ones with children more enjoy their lives in urban cities than those without children. At the same time, the estimates about ages show that the unemployment values have the hump shape. However, the coefficients describe that the age effects are not big. Hence, the older people have lower flow utility values in urban cities and they have higher psychic values for staying at rural areas. This finding is consistent with data observation that migrants are much more younger than the individuals living in rural areas.

In this model, social networks play a key role on two channels: migration cost and job arrival rate. Table 10 displays the role of social networks through two channels. Exogenous model estimates show that the average migration costs for the individuals with networks is is 91.6% of the values for those without networks. The migration costs of endogenous model are larger compared with exogenous model. Also, individuals with networks will pay less around 10% of migration costs. From migration cost equation, it is also found that the married and the individuals with children have larger migration costs. Also, the older cohort individuals have larger migration costs. The important thing to be pointed is that social networks can reduce the search frictions significantly. The arrival rate for the individuals with networks (i.e. $\lambda = 0.19$) is almost twice than those without (i.e. $\lambda = 0.10$) in the exogenous model and there exists similar finding about the endogenous model.

Since I try to maximize the likelihood function, I can use the estimates to simulate individuals' behaviours and to compare the simulated results with data moments. Table 11 and 12 give the comparison between the current version model predicts and data moments. From the comparison, the current estimates fit these moments reasonably well . Next I will discuss the comparison.

Table 11 and 12 give the model fits about earnings and endogenous choices. The estimation method used in this paper is maximized likelihood estimation. The moments reported in Table 11 and 12 are not the targets in the estimation.

Table 11 shows the comparison about the earnings' moments. The column for data gives the selected moments for both migrants' and non-migrants' earnings including the mean and variance of log earnings. The model predicts the earnings for both migrants and non-migrants which fit the data quite well. The model also can capture the earnings with networks are higher than those without networks.

Table 12 gives the model fit for choices. These two models can match the fraction of the individuals with networks and the fraction of networks quite well. Here migrants include the ones who have jobs and the agents with unemployment state in urban cities. If examining the composition of the migrants, we may find that the exogenous model captures the agents with networks will migrate quite well (i.e. 21.86% vs. 21.56%) and under-predict the migration behaviour of individuals without networks (i.e. 7.14% vs. 5.56%). The endogenous model has a little poorer performance to predict the fraction of migrants with networks (21.86% vs. 19.59%) but has a better performance to predict the fraction of migrants without networks. Compared with the real data, the both models under predict the fraction of migrants without networks. Adding unobserved type might help the model to fit the data since it may encourage the high type individuals without networks to migrate and low type individuals with networks prefer to stay in rural areas.

Table 13 gives the counterfactual simulation results for the exogenous model in term of the role of networks (i.e. networks only affect job arrival rate, only affect migration costs, or neither of two channels). The column of model gives the model predicts about choices. The neither column shows the model prediction if social networks do not influence neither job arrival rate and migration costs. It shows that without the effect of social networks, there are only 17.22% of individuals will migrate, which is almost less than 10% compared to the model prediction. From the migration cost column, if social networks only affect migration costs, there are only an additional 2.1% individuals who will migrate. The job arrival rate column shows that if social networks only increase job arrival rates and have no effect on migration costs. There are an additional 9.6% individuals who will migrate. These simulation results show that reducing search frictions is more important compared to reducing migration costs.

Table 14 reports the simulation results for the endogenous model. The neither column shows that if social networks do not affect both channels only 11.48% individuals will migrate. For the exogenous model, this number is 17.22%. Compared with these two models, the endogenous model shows that social networks play an even larger role in terms of migration choices. Since without networks, only 11.5% of people migrate. The model results will tell us that individuals actually effectively use and invest social networks to optimize their lives. When I disentangle the two channels, I get the similar findings as those from the exogenous model. If social networks only affect migration costs, there are additional 2% more people who will migrate. If networks only increase job arrival rate, there are almost 12% more people who will migrate.

6 Conclusion

This paper tries to understand how social networks affect individuals' migration decisions and subsequence labor market outcomes. I construct and structurally estimate a dynamic model of the joint migrantion choices including repeat and return migration, unemployment, social network investment decisions of men and allow agent accumulate their human capital through location specific working experience. In any given period, an individual has to choose between migrating to the city, returning home, labor supply and investment on social networks. In order to distinguish between the two channels through which social networks affect migrations, I allow for the size and presence of the social networks to have a direct effect on migration costs, as well as an indirect effect on labor outcomes via their effect on job arrival rates. I then estimate the model using dataset of Chinese Household Income Projects.

The results show that social networks affect individuals' migration choices and subsequent labor market outcomes through both channels: reduce migration costs and search frictions. Social networks can reduce 10% of migration costs and increase more than 90% of job arrival rates. Compared with these two channels, reducing search frictions play a larger role.

When compared with two structural models, I find that if allow individuals to invest their own social networks, they can effectively response the status of their social networks and try to invest or not to increase or keep their social networks.

To further examine the role of social networks, I will estimate the model with considering network investment choice with unobserved types. Introducing unobserved type would help the model to incorporate the sorting behaviour in terms of ability.

References

Beringhaus, S. and H.G. Seifer-Vogt. 1988. "Temporary versus Permanent Migration: a Decision Theoretical Approach." *Journal of Population Economics* 1:195–211.

Bishop, Kelly. 2012. "A Dynamic Model of Location Choice and Hedonic Valuation." *Working paper* .

Blejer, M.I. and I. Goldberg. 1980. "Return Migration-expectations versus Reality: a Case Study of Western Immigrants to Israel." *Research in population economics* 2.

Borjas, G. and B. Bratsberg. 1996. "Who leaves? The Outmigration of the Foreign-born." *Review of Economics Statistics* 78:165–176.

Borjas, George. 2000. "The Economic Progress of Immigrants." *Issues in the Economics of Immigration* :15–50.

Borjas, G.J. 1999. *The Economic Analysis of Immigration.*, vol. 3A. Amsterdam: North-Holland.

Buchinsky, M, C Gotlibovski, and O Lifshitz. 2012. "Residential Location, Work Location, and Labor Market Outcomes of Immigrants in Israel." *forthcoming Economatrica* .

Burdett, Kiefer N.M. Mortensen D.T., K. and G.R. Neumann. 1984. "Earnings, unemployment, and the allocation of time over time." *Review of Economic Studies* 51(4):559–78.

Carrington, Detragiache, and Vishwanath. 1996. "Migration with Endogenous Moving Costs." *American Economic Review* Vol. 86(4):909–930.

Chiswick, Lee, and Miller. 2005. "Family matters: the role of the family in immigrants' destination language acquisition." *Journal of Population Economics* 18:631–647.

Chiswick, B.R. 1978. "The effect of Americanization on the Earning of Foreign-born Men." *Journal of Political Economy* 86:897–921.

Edin, Fredriksson, and Aslund. 2003. "Ethnic Enclaves And The Economic Success Of Immigrants-Evidence From A Natural Experiment." *The Quarterly Journal of Economics* 118:329–357.

Flinn, C. and J. Heckman. 1982. *Models for the Analysis of Labour Force Dynamics.* Greenwich: JAI Press.

Frijters, P and Bas van der Klaauw. 2006. "Job Search with Nonparticiparion." *The Economic Journal* 116:45–83.

Galor, O and Stark O. 1990. "Migrants' Savings, the Probability of Return Migration, and Migrants." *International Economic Review* 31:463–467.

Gemici, Ahu. 2011. "Family Migration and Labor Market Outcomes." *Working paper* .

Goel, Deepti and Kevin Lang. 2012. "Social Ties and the Job Search of Recent Immigrants." *Working Paper* .

Hare, Denise and Shukai Zhao. 2000. "Lobor Migration as a Rural Development Strategy: a View from the Migration Origin." *Working paper* .

Hare, Densise. 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* 35:45–72.

Harris and M. P. Todaro. 1970. "Migration, unemployment and development: A two sector analysis." *American Economic Review* Vol. 60:126–142.

Heckman, James and B. Singer. 1984. "A Method of Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data." *Econometrica* 52:271–320.

Hu, Zhijun, Chongming Liu, and Zhimin Gong. 2011. "Estimation of the Gini Coefficient in China: 1985-2008." *China Economic Quarterly* 10:1423–1436.

Huang, P. and F.N. Pieke. 2003. "China Migration Country Study."

Jasso, G. and M.R. Rosenzweig. 1982. "Estimating the Emigration Rates of Legal Immigrants Using Administrative and Survey Data: the 1971 Cohort of Immigrants to the US." *Demography* 19:279–290.

Keane and K. Wolpin. 1994. "The Solution and Estimation of Discrete Choice Dynamic Programming Models by Simulation and Interpolation: Monte Carlo Evidence." *The Review of Economics and Statistics* Vol. 76 (4):648–672.

———. 1997. "The Career Decisions of Young Men." *Journal of Political Economy* Vol. 105(3):473–522.

Kennan and Walker. 2011. "The Effect of Expected Income on Individual Migration Decisions." *Econometrica* Vol. 79:211–251.

Kono. 2006. "Employment with connections: Negative network effects." *Journal of Development Economics* Vol. 81 (1):244–258.

Lessem, Rebecca. 2011. "Labor Migration and Wage Growth in Malaysia." *Working paper* .

Lin, F. Cai, Justin and Z. Li. 1996. *The China Miracle: Development Strategy and Economic Reform*. Hong Kong: Chinese University Press.

Machin, S. and A. Manning. 1999. *Handbook of Labor Economics*, vol. 3C, chap. The Cause and Consequences of Longterm Unemployment in Europe. Amsterdam: North-Holland, 3085–139.

McCall, B. 1996. "Unemployment Insurance Rules, Joblessness, and Part-time Work." *Econometrica* 64:647–682.

Meng, Xin. 2000. *Labout Market Reform in China*. Cambridge University Press.

Mortensen, D.T. 1986. "Job Search and Labor Market Analysis." *Handbook of Labor Economics* 2:849–919.

Mortensen, D.T. and G.R. Neuman. 1984. *Choice or Chance? A Structural Interpretation of Individual Labor Market Dynamics*. Berlin: Springer-Verlag, 98–131.

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 Jiangxu, China." *Journal of International Development* 11:662–672.

Thom, Kevin. 2010. "Repeated Circular Migration: Theory and Evidence from Undocumented Migrants." *Working paper* .

Todaro, M. P. 1969. "A Model of Labor Migration and Urban Unemployment in Less Development Countries." *American Economic Review* 59:138–48.

Uhlendorff, A. and K.F. Zimmermann. 2006. "Unemployment dynamics among Migrants and Natives." *Working paper* .

Van den Berg, G. 2001. "Duaration Models: Specification, identication, and multiple duration." *Handbook of Econometrics* 5:3381–460.

Van den Berg, G.J. 1990. "Search Behavior, Transitions to Nonparticiparion and the Duration of Unemployment." *Economic Journal* 100:842–65.

Weiner, S.E. 1984. *A Survival Analysis of Adult Male Black-White Labor Market Flows*. Berlin: Springer-Verlag, 132–57.

Wolpin, K.I. 1987. "Estimating a Structural Search Model: the Transition From School to Work." *Econometrica* 55(4):801–17.

Wu, Harry Xiaoying. 1994. "Rural to Urban Migration in the People's Republic of China." *China Quarterly* 139:669–98.

Yao, Y. 2001. "Egalitarian Land Distribution and Labor Migration in Rural China." *working paper* .

Zhang, Junfu and Zhong Zhao. 2011. "Social-Family Network and Self- Employment: Evidence from Temporary Rural-Urban Migrants in China." *working paper* .

Zhao, Yaohui. 1997. "Labor Migration and Returns to Rural Education in China." *American Journal of Agricultural Economics* 79:1278–87.

———. 2000. "Rural-to-Urban Labor Migration in China: the Past and the Present." .

———. 2002. "Causes and Consequences of Return Migration: Recent Evidence from China." *Journal of Comparative Economics* 30:376–394.

———. 2003. "The Role of Migrant Networks in Labour Migration: the Case of China." *Contemporary Economic Policy* 21(4):500–511.

Zhao, Zhong. 2005. "Migration, Labor Market Flexibility, and Wage Determination in China: A Review." *Developing Economics* XLIII-2:285–312.

7 Appendix

Figure 1: Sample Cities in China

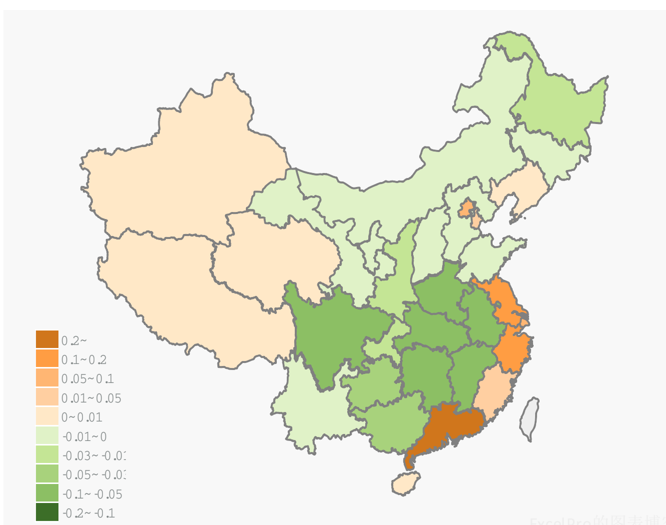


Figure 2: Migration flows in China From 2000 to 2005

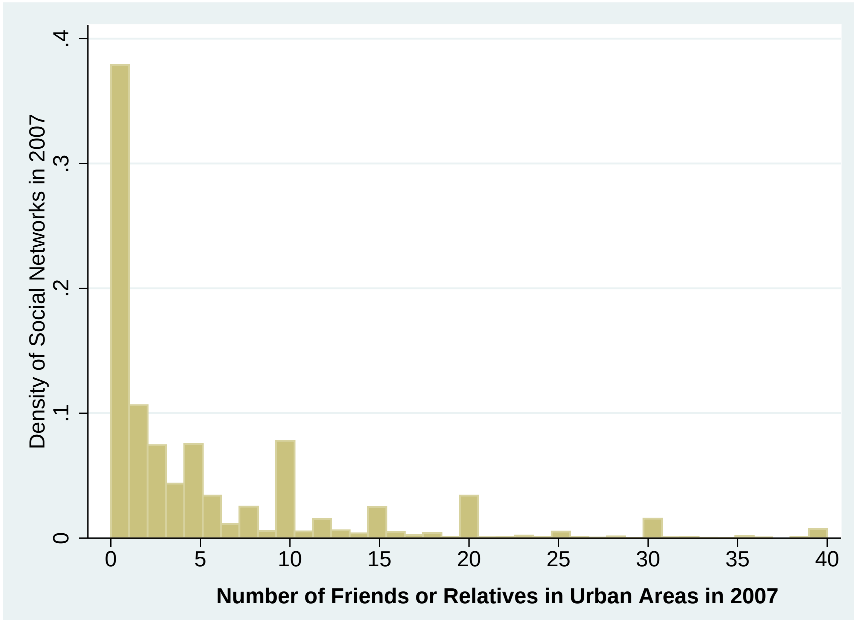


Figure 3: The Size of Social Networks in 2007

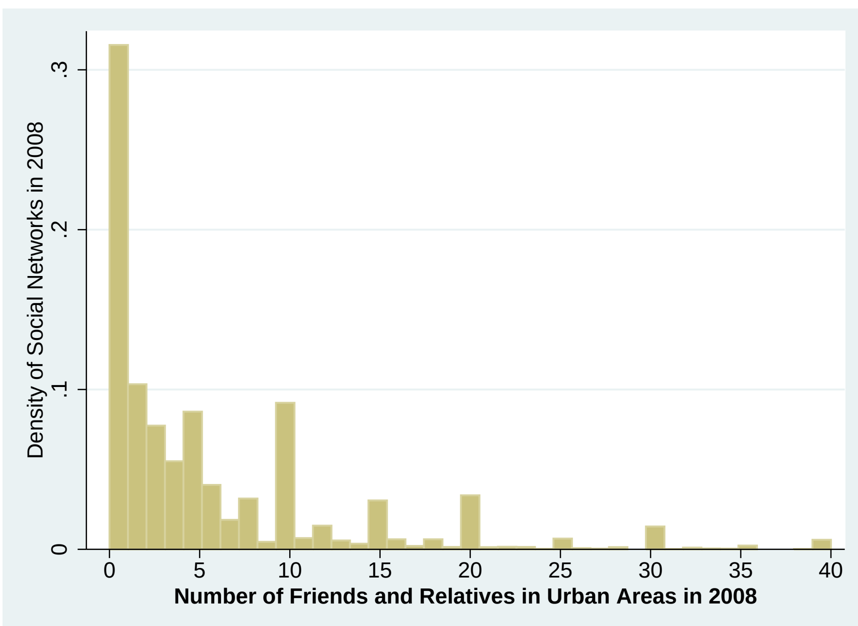


Figure 4: The Size of Social Networks in 2008

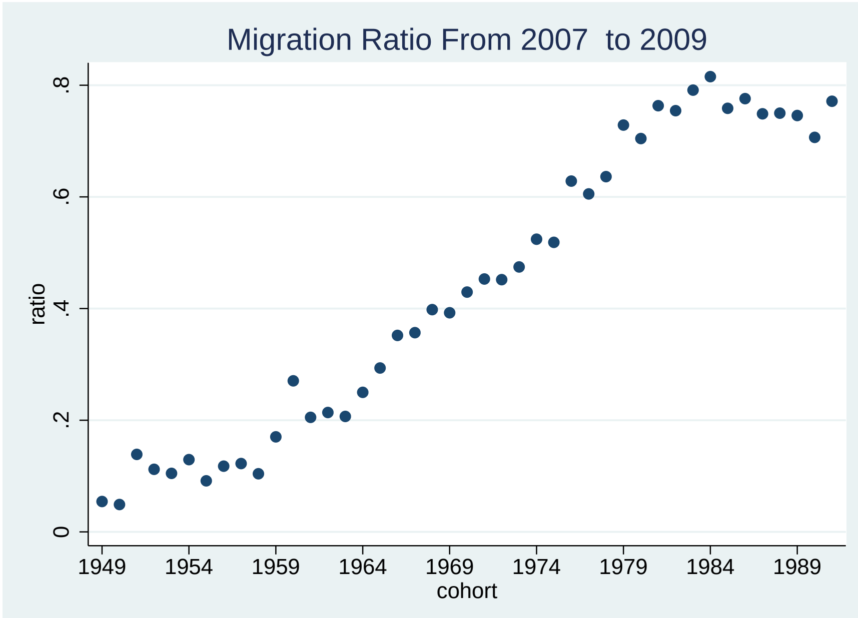


Figure 5: Migration Fraction in Different Cohorts

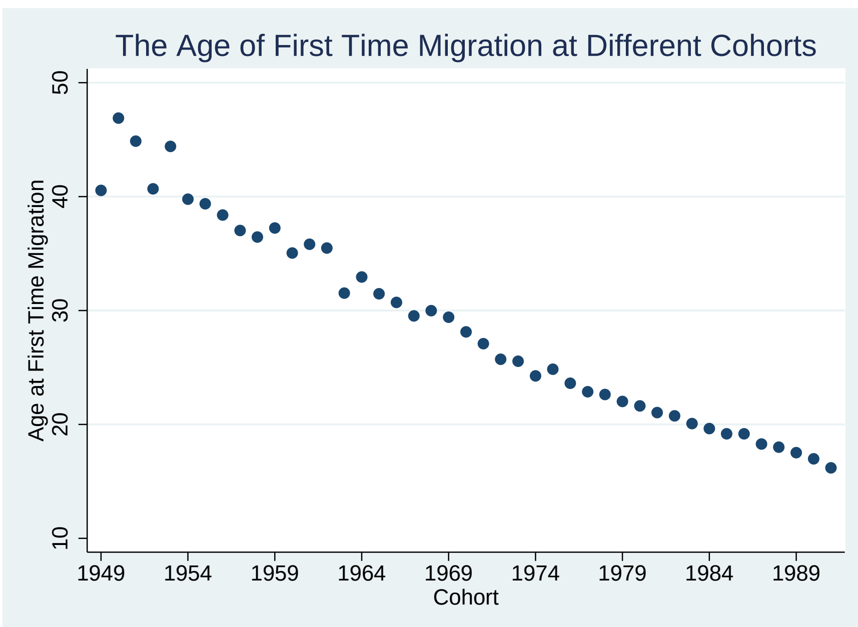
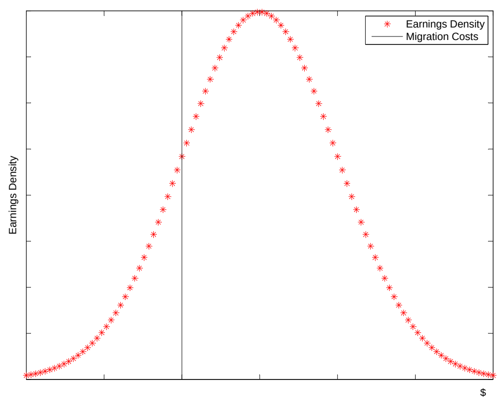
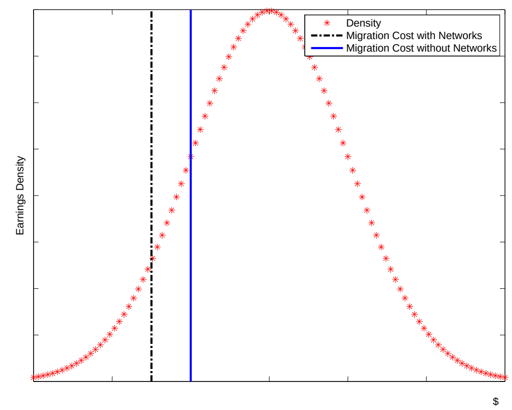


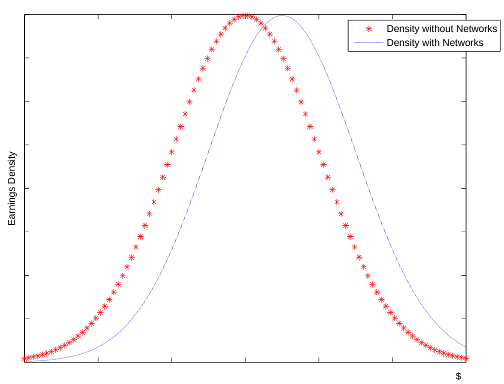
Figure 6: Age at the First Time Migration in Different Cohorts



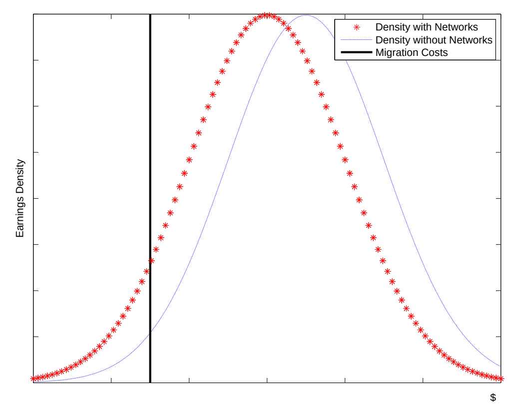
(a) With the Friction: Migration Costs



(b) Social Networks Affect Migration Costs



(c) With the Friction: Job Arrival Rate



(d) Social Networks Affect Job Arrival Rate

Figure 7: How Social Networks Affect Migrants' Earnings

Table 1: Interprovincial Migration in China, 1990-2005(In Thousands)

	1990-1995	migration				1995-2000		migration				2000-2005		migration			
		In	Out	Net	Net(%)			In	Out	Net	Net(%)			In	Out	Net	Net(%)
1	Guangdong	1947	221	1726	16.2	1	Guangdong	11501	438	11063	34.3	1	Guangdong	11996	1715	10281	27
2	Shanghai	726	122	604	5.7	2	Shanghai	2168	163	2005	6.2	2	Zhejiang	5062	1041	4021	10.6
3	Beijing	694	117	577	5.4	3	Zhejiang	2715	970	1745	5.4	3	Shanghai	3025	375	2650	7
4	Jiangsu	969	450	519	4.9	4	Beijing	1890	174	1715	5.3	4	Jiangsu	3290	1328	1963	5.2
5	Xinjiang	566	150	416	3.9	5	Xinjiang	1142	217	925	2.9	5	Beijing	2246	330	1916	5
6	Liaoning	435	197	239	2.2	6	Fujian	1346	625	722	2.2	6	Fujian	1934	802	1132	3
7	Tianjin	223	62	161	1.5	7	Jiangsu	1908	1241	667	2.1	7	Tianjin	908	107	802	2.1
8	Shandong	527	382	145	1.4	8	Tianjin	492	104	388	1.2	8	Xinjiang	577	182	395	1
9	Fujian	344	220	125	1.2	9	Liaoning	755	380	375	1.2	9	Liaoning	674	416	257	0.7
10	Hebei	503	417	87	0.8	10	Yunnan	733	398	335	1	10	Hainan	191	158	33	0.1
11	NeiMongol	275	249	27	0.3	11	Hainan	218	130	88	0.3	11	Ningxia	74	68	7	0
12	Shanxi	158	140	18	0.2	12	Shanxi	383	334	49	0.2	12	Tibet	26	31	-6	0
13	Tibet	38	28	10	0.1	13	Ningxia	129	87	41	0.1	13	Qinghai	74	85	-12	0
14	Hainan	104	102	2	0	14	Tibet	71	35	35	0.1	14	NeiMongol	394	417	-23	-0.1
15	Ningxia	49	54	-6	-0.1	15	Shandong	904	878	26	0.1	15	Yunnan	469	601	-132	-0.3
16	Qinghai	51	77	-25	-0.2	16	Qinghai	77	123	-46	-0.1	16	Shanxi	210	345	-135	-0.4
17	Yunnan	207	242	-35	-0.3	17	Hebei	770	872	-102	-0.3	17	Shandong	924	1123	-199	-0.5
18	Zhejiang	466	514	-49	-0.5	18	NeiMongol	325	441	-116	-0.4	18	Jilin	218	532	-315	-0.8
19	Shaanxi	163	265	-101	-1	19	Jilin	254	529	-275	-0.9	19	Gansu	118	494	-376	-1
20	Hubei	271	382	-111	-1	20	Shaanxi	423	719	-296	-0.9	20	Hebei	612	990	-378	-1
21	Gansu	140	251	-112	-1	21	Gansu	204	561	-357	-1.1	21	Shaanxi	255	827	-572	-1.5
22	Jilin	150	295	-145	-1.4	22	Heilongjiang	301	940	-639	-2	22	Heilongjiang	195	1020	-825	-2.2
23	Guizhou	152	402	-250	-2.3	23	Chongqing	448	1103	-655	-2	23	Chongqing	427	1437	-1010	-2.7
24	Jiangxi	125	514	-389	-3.6	24	Guizhou	261	1232	-970	-3	24	Guizhou	531	1766	-1235	-3.2
25	Heilongjiang	224	614	-389	-3.7	25	Guangxi	287	1838	-1551	-4.8	25	Guangxi	397	2123	-1726	-4.5
26	Guangxi	120	554	-434	-4.1	26	Hubei	606	2210	-1604	-5	26	Jiangxi	499	2476	-1977	-5.2
27	Henan	270	740	-470	-4.4	27	Henan	470	2309	-1839	-5.7	27	Hubei	501	2715	-2214	-5.8
28	Hunan	215	704	-489	-4.6	28	Jiangxi	236	2681	-2445	-7.6	28	Hunan	501	3328	-2827	-7.4
29	Anhui	155	744	-589	-5.5	29	Anhui	313	2893	-2579	-8	29	Henan	280	3433	-3154	-8.3
30	Sichuan*	395	1457	-1062	-10	30	Hunan	363	3261	-2899	-9	30	Anhui	671	3836	-3165	-8.3
						31	Sichuan	590	4396	-3806	-11.8	31	Sichuan	763	3941	-3178	-8.4
		9189	9189.00	0				32282	32282	0				38042	38042	0	

Before 2000, Chongqing was part of Sichuan province. The data for Sichuan from 1990 to 1995 includes Chongqing.
Source from NPSSO(1997), SC and NBS (2002,2007)
Net%=Net migration/Total national migration×100%.

Table 2

Summary Statistics		
	male	observation
network size	6.83	9256
	(12.61)	
network category		9256
0	31.81%	
1 ($1 \leq$ sn ≤ 6)	29.39%	
2 ($6 <$ sn ≤ 11)	12.52%	
3 ($11 <$ sn ≤ 21)	9.76%	
4 (>21)	6.52%	
network invest 07	61.26%	8026
	(0.49)	
network invest 08	77.76%	7564
	(0.42)	
education year	8.28	9256
	(2.14)	
education level		9256
primary or less	18.34%	
middle school	63.49%	
high school	14.55%	
college or above	3.62%	
urban earnings in 2007	1343.43	5325
	(1731.09)	
urban earnings in 2008	1345.82	5532
	(1427.93)	
urban earnings in 2009	1598.15	5415
	(1710.65)	
rural earnings in 2007	275.32	
	(377.58)	
rural earnings in 2008	298.17	
	(439.21)	
rural earnings in 2009	299.81	
	(498.95)	

1. Sn: the size of social networks.
2. Earnings have been adjusted by location price index and cpi price index.
3. Earnings are in Chinese currency yuan which is closed to 1/6 of 1 U.S. Dollar.
4. Numbers with parentheses are standard deviations

Table 3

Summary Statistics			
	Total	Migrants	Non-migrants
Education	8.28	8.63	8.01
	(2.14)	(1.92)	(2.25)
With Networks	8.37	8.73	8.14
	(2.12)	(1.93)	(2.22)
Without Networks	8.03	8.38	7.73
	(2.16)	(1.89)	(2.27)
Age	38.86	31.00	42.06
	(11.93)	(9.16)	(11.44)
With Networks	39.01	31.01	42.48
	(11.96)	(9.12)	(11.37)
Without Networks	38.40	30.97	40.90
	(11.84)	(9.28)	(11.55)
Marriage	0.78	0.61	0.85
	(0.41)	(0.49)	(0.35)
With Networks	0.79	0.61	0.86
	(0.41)	(0.49)	(0.35)
Without Networks	0.78	0.61	0.83
	(0.42)	(0.49)	(0.37)
Urban Earnings	1404.67	1404.67	
	(1168.82)	(1168.82)	
With Networks	1431.78	1431.78	
	(1232.21)	(1232.21)	
Without Networks	1310.47	1310.47	
	(909.06)	(909.06)	
Rural Earnings	282.63		282.63
	(477.38)		(477.38)
With Networks	284.96		284.96
	(480.50)		(480.50)
Without Networks	276.12		276.12
	(468.48)		(468.48)
Job Search Duration	1.90	1.90	
	(3.08)	(3.08)	
With Networks	1.92	1.92	
	(3.09)	(3.09)	
Without Networks	1.85	1.85	
	(3.06)	(3.06)	

1. Numbers with parentheses are standard deviations
2. Earnings have been adjusted by location price index and cpi price index
3. Job search period unit is monthly

Table 4

The Mobility of the Size of Social Networks from 2008 to 2009					
A. If agents invested social networks in 2008					
	The size of social networks in 2008				
in 2009	0	1	2	3	4
0	43.03%	13.07%	10.01%	10.32%	12.83%
1	36.99%	57.16%	30.86%	24.87%	19.71%
2	10.88%	14.83%	32.65%	21.15%	15.91%
3	4.19%	8.79%	19.20%	31.13%	20.43%
4	4.92%	6.15%	7.27%	12.52%	31.12%
B. If agents did not invest social networks in 2008					
in 2009	0	1	2	3	4
0	41.32%	18.65%	10.30%	8.95%	6.85%
1	39.85%	56.95%	39.09%	29.18%	24.66%
2	7.44%	14.06%	30.30%	26.07%	18.49%
3	4.13%	6.45%	15.15%	25.29%	23.29%
4	7.25%	3.89%	5.15%	10.51%	26.71%
The difference between A and B					
2009	0	1	2	3	4
0	1.71%	-5.57%	-0.29%	1.37%	5.98%
1	-2.87%	0.20%	-8.23%	-4.31%	-4.94%
2	3.44%	0.77%	2.34%	-4.92%	-2.58%
3	0.06%	2.34%	4.05%	5.84%	-2.86%
4	-2.34%	2.26%	2.12%	2.02%	4.40%

1. column is the category of social networks in 2008
2. Part A is that the probability of network transition matrix from 2008 to 2009 if agents did networks investment in 2008
3. Part B provides the information if agents did not invest their networks in 2008
4. Part C gives the difference between Part A and Part B

Table 5

Ordered probit model		
	(1)	(2)
investment	0.2587***	0.0310
education		
2	0.1029***	0.0353
3	0.1879***	0.0463
4	0.0761***	0.0728
sn(t-1)		
1	0.5355***	0.0313
2	1.0730***	0.0429
3	1.3338***	0.0467
4	1.5447***	0.0556
cut-off		
1	0.0973	0.0674
2	1.4230***	0.0690
3	2.0269***	0.0702
4	2.7791***	0.0731
No. obs.	7359	

1. dependent variable: the size of social networks in current period
2. sn(t-1): the size of social networks in last period
3. column 1: estimates, column 2: standard errors
4. $*$ significant at 10% level, $**$ at 5% level and $***$ at 1% level

Table 6

	Probit	OLS
Network Size		
$1(1 \leq sn \leq 6)$	0.0240	0.0190
	(0.023)	(0.014)
$2(6 < sn \leq 11)$	0.0810	0.0240
	(0.030)	(0.017)
$3(11 < sn \leq 21)$	0.0619	0.0805
	(0.034)	(0.019)
$4(sn > 21)$	0.0927	0.0934
	(0.036)	(0.022)
Education		
Middle school	0.0296	0.0651
	(0.030)	(0.024)
High school	-0.0800	0.1486
	(0.039)	(0.028)
College or above	-0.0340	0.1455
	(0.058)	(0.044)
Marriage	0.0744	0.0363
	(0.036)	(0.020)
Num of Kids	-0.0042	-0.0425
	(0.015)	(0.011)
Age	-0.0029	0.0055
	(0.001)	(0.000)
Num of Obs	9060	

1. The first column is the estimation results of probit model for migration choice
2. The results in column (1) are average marginal effects
3. In Column (2), the dependent variable is urban earnings for migrants with jobs
4. Standard errors are in parentheses
5. Sn means the size of Networks

Table 7: Survival Analysis of the Age of First Time Migration

	Coefficient	Standard Error
Education	-0.0016	0.0058
Cohort		
(1960-1969)	-0.5852	0.0364
(1970-1979)	-1.4855	0.0387
(1980-1991)	-2.4451	0.0350
Year		
(1984-1991)	0.1335	0.7340
(1992-2000)	0.1299	0.0842
(2001-2009)	-0.1654	0.0348
Constant	4.1304	0.0544
γ	0.5718	0.0077

1. The first column is the estimation results of survival regression with loglogistics distribution
2. The second column is the standard error
3. The omitted cohort dummy is the cohort (1949-1959)
4. The omitted cohort year dummy is the period before 1984 which is the period that rural-urban migration are prohibited

Table 8: Estimation Results for Exogenous Model

Earning Equation(Urban)		Social Network Probit Equation	
edu year	0.0143	marriage	0.0883
expu	0.0040	num of children	-0.0364
expr	0.0035	sn_{t-1}	4.2777
$expu^2 \times 100$	-0.0018	constant	-1.9191
$expr^2 \times 100$	0.0001	Psychic value	
constant	6.4450	age	0.0369
Earning Equation(Rural)		$age^2 \times 100$	-0.0018
edu year	0.0087	marriage	0.0013
expu	0.0018	num of children	-0.0870
expr	0.0057	constant	0.0854
$expu^2 \times 100$	-0.0037	Job arrival rate	
$expr^2 \times 100$	-0.0005	social network	0.7409
constant	4.1083	edu year	0.0028
Unemployment value		constant	-2.2334
marriage	0.1068	Job separation rate	
num of children	0.7443	edu year	0.0005
age	0.0014	constant	-3.9982
$age^2 \times 100$	-0.0354	Return Migration Cost	
constant	0.0276	marriage	0.8663
Migration cost		num of children	0.0798
social network	-1.6364	cohort	-0.0034
marriage	0.1534	constant	8.7035
num of children	0.0395	loglike	-37753.67
cohort	-0.1378		
constant	15.5154		

1: cohort is the year of birth-1948

2: social network is indicator variable.

3: experiences are monthly experience and age is annual age.

Table 9: Estimation Results for Endogenous Model

Earning Equation(Urban)		Social Network Probit Equation	
edu year	0.0243	marriage	0.0890
expu	0.0040	num of children	-0.0361
expr	0.0035	sn_{t-1}	3.7457
$expu^2 \times 100$	-0.0018	inv_{t-1}	0.0907
$expr^2 \times 100$	0.0005	constant	-1.6841
constant	6.3050	Psychic value	
Earning Equation(Rural)		age	0.0359
edu year	0.0087	$age^2 \times 100$	-0.0014
expu	0.0036	marriage	0.0013
expr	0.0075	num of children	-0.0870
$expu^2 \times 100$	-0.0037	constant	0.0853
$expr^2 \times 100$	-0.0005	Job arrival rate	
constant	3.9723	social network	0.7349
Unemployment value		edu year	0.0028
marriage	0.1068	constant	-2.2334
num of children	0.7443	Job separation rate	
age	0.0014	edu year	0.0005
$age^2 \times 100$	-0.0004	constant	-3.9982
constant	0.0256	Return Migration Cost	
Migration cost		marriage	0.8663
social network	-1.6367	num of children	0.0798
marriage	0.1536	cohort	-0.0034
num of children	0.0393	constant	8.7035
cohort	-0.1448	loglike	-26094.87
constant	17.5154		

1: cohort is the year of birth-1948

2: social network is indicator variable.

3: experiences are monthly experience and age is annual age.

Table 10: The impact of Social Networks

	Exogenous		Endogenous	
	Migration Cost	Job Arrival Rate	Migration Cost	Job Arrival Rate
Social networks coefficient	-1.64	0.74	-1.63	0.73
Average				
With Networks	91.64%	0.19	106%	0.18
Without Network	100%	0.10	119.91%	0.10

Table 11: Model Fit: Earnings

	Data	Exogenous	Endogenous
migrants			
log(earnings)	7.1068	7.1447	7.1168
$sd(log(earnings))$	0.2681	0.2504	0.3755
with networks	7.1229	7.1467	7.1190
$sd(log(earnings))$	0.2695	0.2497	0.3767
without networks	7.0580	7.1365	7.1090
$sd(log(earnings))$	0.2612	0.2528	0.3696
non-migrants			
log(earning)	5.0310	5.0205	5.0070
$sd(log(earnings))$	1.6292	1.6303	1.8262
with networks	5.0677	5.0334	5.0277
$sd(earnings)$	1.5418	1.6279	1.8213
without networks	4.9410	4.9908	4.9581
$sd(earnings)$	1.8318	1.6346	1.8337

1. Migrants include people who were born in rural areas and resided in urban cities who can be employed or unemployed.

Table 12: Model Fits: Choices

	Data	Exogenous	Endogenous
social networks	72.30%	72.38%	71.99%
*migrants**	29.00%	27.12%	25.59%
with networks	21.86%	21.56%	19.59%
without networks	7.14%	5.56%	6.00%
return migrants	0.67%	1.06%	0.77%
with networks	0.45%	0.80%	0.57%
without networks	0.22%	0.26%	0.20%
moving (rural to urban)	0.76%	1.18%	0.82%
with networks	0.54%	0.97%	0.69%
without networks	0.22%	0.22%	0.12%
job search duration	1.91	2.34	2.66

1. Migrants include people who were born in rural areas and resided in urban cities who can be employed or unemployed.

Table 13: Counterfactual Results: Exogenous Social Networks

			Social Networks just Affect	
	Exogenous	Neither	Migration Cost	Job Arrival Rate
social networks	72.38%	72.38%	72.38%	72.38%
migrants	27.12%	17.22%	19.37%	26.82%
with networks	21.56%	12.06%	14.29%	20.76%
without networks	5.56%	5.16%	5.08%	6.06%
non-migrants	72.88%	82.78%	80.63%	73.18%
with networks	49.27%	60.31%	58.08%	51.61%
without networks	23.61%	22.47%	22.55%	21.43%
job search duration	2.34	2.68	2.55	2.51

1. Migrants include people who were born in rural areas and resided in urban cities who can be employed or unemployed.

Table 14: Counterfactual Results: Endogenous Social Networks

	Data	Endogenous	Neither	Social Networks just Affect Migration Cost	Social Networks just Affect Job Arrival Rate
social networks	72.30%	71.99%	71.99%	71.99%	71.99%
migrants	29.00%	25.59%	11.48%	13.43%	23.27%
with networks	21.86%	19.59%	8.15%	9.85%	17.48%
without networks	7.14%	6.00%	3.34%	3.58%	5.79%
non-migrants	71.00%	74.41%	88.52%	86.57%	76.73%
with networks	50.44%	52.40%	63.84%	63.14%	54.51%
without networks	20.56%	22.01%	24.67%	24.43%	22.22%
job search duration	1.91	2.66	2.94	2.79	2.8

1. Migrants include people who were born in rural areas and resided in urban cities who can be employed or unemployed.

Table 15: Policy Simulation Results

	Exogenous Model		Endogenous Model	
	Policy (Yuan)	Government Budget (Billion Yuan)	Policy (Yuan)	Government Budget (Billion Yuan)
Unemployment Benefit	500	9.51	943	24.86
Migration Cost	421	1.87	703	3.77
Job Arrival Rate	0.14		0.30	