Social Networks as Investment: Rural-Urban Migration in China*

Jin Zhou

Department of Economics and
Center for the Economics of Human Development
University of Chicago
December 1, 2021

*I am grateful to Salvador Navarro, Nirav Mehta, Audra Bowlus, and Terry Sicular for their guidance. I thank James Heckman, Steven Durlauf, Lance Lochner, Joseph Altonji, Jaap Abbring, Hanming Fang, Paul Sullivan, John Ham, Chris Robinson, Todd Stinebrickner, and David Rivers for their helpful comments. I also thank seminar participants at the University of Western Ontario and conference participants at the Human Capital and Economic Opportunity on migration conference, the Society of Labor Economists World Meeting, and the UM-MSU-UWO Labor Day Conference. Fanmei Xia contributed highly competent and insightful research assistance. This research would not have been possible without financial support from the Social Sciences and Humanities Research Council of Canada. An online appendix can be found at https://www.dropbox.com/s/03muhkvawvv7ban/Network_Zhou_Appx.pdf?dl=0.

Abstract

Numerous empirical studies have documented a strong association between an individual's social network and his migration decisions. To understand how social networks affect labor market outcomes and migration decisions and explain the conflicting findings in the current literature, I develop and estimate a dynamic model with return and repeated migration, the dynamics of social network that evolves depending on individuals' network investment, and labor market transitions. The model distinguishes between two channels through which social networks may affect migration decisions: (1) a direct effect on migration costs and (2) an indirect effect on labor market outcomes through the job arrival rate. I use the model to study one of the largest ongoing internal migrations in human history: rural-urban migration in China. I use panel data from the Chinese Household Income Project (2007-2009) to estimate the model. The estimation results show that social networks affect both channels significantly. Individuals with networks have 40% higher job arrival rates than those without networks, on average. In addition, social networks reduce average migration costs by 7%. I also show that policies that directly lower migration costs may be cost-effective at increasing rural-urban migration in China. These policy experiments also show that government policies crowd out individual social network investment. Hence, the policy which keeps the investment incentive can achieve the same government target with a lower cost.

JEL Codes: O1, J6, C5, D8

Keywords: Internal Migration, Job Search, Social Networks.

1 Introduction

A strong association between social networks[1] and migration decisions has been consistently documented in numerous empirical studies. Some researchers point out that individuals with social networks in destination places are more likely to migrate (e.g., Munshi, 2003). Social networks are often viewed as an important non-market factor through which individuals reduce their labor market frictions and improve their labor market outcomes. However, there are conflicting findings of the quantitative effects of social networks on the labor market outcomes of migrants. For example, social networks may provide access to better jobs (e.g., Munshi, 2003 and Edin, Fredriksson, and Åslund, 2003) or to less desirable ones (e.g., Borjas, 2000 and Barry R. Chiswick and Miller, 2005). These conflicting findings indicate that we should investigate the mechanisms of the role of social networks in the labor market.

To formally understand how the social networks affect the individuals' labor market outcomes, I develop a dynamic model in which social networks affect an individual's migration decision and labor market outcomes. To capture the key feature of rural-urban migration in China, the model allows for a return and repeated migration and labor market frictions (e.g., migration cost and information frictions). Following the existing migration literature, there are 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), decreasing individuals' migration reservation values causing individuals with networks to be more likely to migrate. Second, social networks provide information about labor markets and then increase the probability of getting job offers in the destinations (e.g., Montgomery, 1991; Kono, 2006; Goel and Lang, 2019; Buchinsky, Gotlibovski, and Lifshitz, 2014). Under both of these mechanisms, individuals with social networks are more likely to migrate.

Although both of these mechanisms can explain why individuals with networks are more likely to migrate, they can potentially be distinguished as they have different implications for migrants' earnings. On the one hand, individuals with social networks have lower migration costs which cause lower reservation earnings. Under this situation, migrants with networks are more likely to take lower-paying jobs than sim-

[1] In this paper, the social networks indicate the friends or relatives at the potential migration destinations, not rural risk-sharing networks, which are well documented in the recent literature. In China, the rural households usually smooth their consumption by precautionary saving, i.e., Giles and Yoo (2007). Therefore, the activities of risk sharing across households (e.g., Munshi and Rosenzweig (2016); Morten (2019)) are not common in China.

ilar individuals without networks. On the other hand, social networks reduce search frictions, for example, by increasing the job arrival rate and hence their reservation wage. In this case, migrants with networks will have higher earnings than similar individuals without networks. These different implications for migrants' earnings may be one reason to explain the conflict findings of the impact of social networks on earnings. One of the goals of this paper is to quantify the effects through different channels that social networks may play with regard to labor market outcomes and migration decisions.

Identification of these effects is complicated by the fact that social networks are unlikely to arise independently of individuals' labor market prospects. That is, individuals make investment choices in their social networks by comparing the loss from the investment in the network to the benefit from increasing the probability of having a social network. In this paper, I account for this possibility by formally modeling individuals' social network investment decisions, which benefits our understanding of how individuals respond to market frictions through their social networks. Considering social network investment decisions also helps to evaluate potential government migration policies. The effects of government policies on market frictions and migration costs are likely to result in differential responses by individuals in terms of their social investment decisions and, ultimately, their migration outcomes. Failure to account for these feedback effects may lead to inaccurate policy evaluation.

Understanding all these different channels through which social networks operate is crucial for accurately designing migration policies. For example, the Chinese government aims to increase the urbanization rate to 60% by 2020 in New Urbanization Plan (2014–2020).[2] Whether social networks are substitutes or complements to government policies aimed to increase migration may significantly affect their cost effectiveness.[3]

Besides accounting for the impact of social networks, the model in this paper also contains a number of mechanisms through which individuals' migration decisions are affected. First, I allow individuals to accumulate human capital within a search framework. Individuals' earnings reflect both their observed characteristics (e.g., education), their location-specific human capital accumulation (i.e., urban and rural), and unobserved endowments (i.e., ability).

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

[2]See http://www.gov.cn/zhengce/2014-03/16/content_2640075.htm.

[3]The situation in China is more complex than what is assumed in the model. Policies are implemented at both local and national government levels, and there exist some barriers to migration that are not formally included in the model. Despite the shortcomings, the model is able to showcase the importance of non-market institutions in developing countries (e.g., China).

market. Individuals 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. This setting incorporates one of the main features of rural-urban migration in China: most people do not migrate permanently.[4]

To study the role of social networks, I examine one of the largest migration episodes of the 20th century: rural to urban migration in China. The current internal rural-urban migration in China provides an ideal setting to examine the role of social networks in a labor market with frictions. Hare and Zhao (2000), Meng (2000) and Zhao (2003) show that social networks are strongly correlated with rural-urban migration in China. Zhang and Zhao (2015) find that social networks also affect migrants' subsequent labor market outcomes. However, these papers do not distinguish the social network effects through the two different channels (i.e., lowering migration costs and reducing search frictions) discussed above.

I estimate the model using the Chinese Household Income Project (CHIP) data. The estimation results show that social networks both significantly reduce migration costs and increase the job arrival rate. The job arrival rate for individuals with networks is 40% higher compared to those without networks. Social networks reduce migration costs by 7% on average. To analyse the importance of these two channels, I simulate the model and show that migration decisions are affected more by the impact of social networks on reducing search frictions than by the impact on reducing migration costs. If I shut down the effects of social networks on both channels, only 15% of rural people migrate. Allowing social networks only to affect migration costs leads to 17% of rural people migrating. If social networks only increase the job arrival rate, 27% of rural people will migrate, compared to 29% in the data.

The simulation results also illustrate how individuals respond to the impact of social networks through network investment. When social networks affect both channels, 63% of individuals invest in their social networks. If social networks only lower migration costs, the fraction of individuals who invest decreases to 47%. When social networks only affect the job arrival rate, 56% of individuals invest in their social networks. The results also show that individuals who invest in their social networks are more likely to be the ones living in rural areas and the ones unemployed in urban areas.[5]

I simulate three different policies to achieve the stated Chinese government's goal of a 60% urbanization rate (i.e., (a) an unconditional lump-sum subsidy for

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

[5]In the paper, the investment of social networks is defined as the activities of sending gifts, keeping in contact with the friends or relatives.

rural individuals who migrate, (b) the provision of unemployment benefits for rural migrants in urban areas, and (c) a migration cost subsidy for rural people targeted only to those who have social networks in urban areas). The simulation results show that the policy of conditional lump-sum transfers for migrants with networks would cost less than the other two policies.

In order to demonstrate the importance of accounting for individuals' decisions to invest in social networks, I also compare the effects of these policies to those obtained in a model estimated under the restriction that individuals are not allowed to invest in their social networks. I find that the government has to spend substantially more if this were the case. Moreover, since individuals cannot use network investment to increase or keep their network status, they have to pay higher migration costs and (or) face more search frictions in urban areas if they migrate. As a consequence, a researcher that ignores social network investment would estimate that, to achieve the same urbanization rate, the government will have to spend more to offset the effect of no investment in social networks.

The rest of the paper is organized as follows. Section 2 provides a review of the relevant literature. Section 3 presents background on rural-urban migration in China, describes the data in detail, and provides a preliminary empirical examination of the key mechanisms in the model. In Section 4, the model and identification conditions is discussed. I also present the estimation procedure, including challenges and solutions. Estimation results and counter-factual simulations are presented in Section 5. Section 6 concludes.

2 Literature Review

The existing migration literature has two main findings on the role of social networks. First, individuals with social networks at destinations are more likely to migrate (i.e., Munshi, 2003). Although Munshi and Rosenzweig (2016) shows how caste-based rural insurance networks have negative effects on rural male migration decisions in India, Hare and Zhao (2000), Meng (2000) and Zhao (2003) find that social networks in urban areas are also positively correlated with rural-urban migration in China. Second, social networks affect migrants' labor market outcomes. For example, social networks may provide access to better jobs (e.g., Munshi, 2003 and Edin, Fredriksson, and Åslund, 2003) or to less desirable ones (e.g., Borjas, 2000 and Barry R. Chiswick and Miller, 2005). Zhang and Zhao (2015) examined the correlation between social-family networks and rural migrants' self-employment in China. They find social-family networks increase migrants' employment probabilities in urban areas.

Recent literature documented that the endogenous relationships between the migration decisions and risk sharing across households (e.g., Munshi and Rosenzweig, 2016; Morten, 2019; Munshi, 2020; and Meghir et al., 2020), which extends the standard migration model considering the income differentials between the locations (e.g., Lewis, 1954; Sjaastad, 1962; Harris and Todaro, 1970). In China, the role of social networks is mostly presented as sharing the labor information or providing monetary support at the destinations.[6] The key contribution of this paper is to examine how the social networks affect the rural individuals' migration decisions and their labor market outcomes through different channels.

Despite numerous empirical studies, it is not clear whether social networks positively affect migrants' earnings. For example, social networks may provide access to better jobs or less desirable ones. There is an ambiguous effect on earnings because social networks affect individuals' earnings through different channels: migration costs and search frictions. The net effect of social networks may vary in different economic environments.

Carrington, Detragiache, and Vishwanath (1996) consider the role of social networks on migration behavior through migration costs. They built a dynamic model to analyze why more black people migrated from the South to the North during the U.S. Great Migration period even though they faced a smaller wage gap. They show that social networks can influence individuals' migration decisions since they may 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 the same social network. Their model does not distinguish search frictions from migration costs in their model either.

Besides the friction of existing migration costs, search frictions in the destination labor markets also affect individuals' migration decisions. Buchinsky, Gotlibovski, and Lifshitz (2014) examine the effect of a few alternative national migration policies on the regional location choices and labor market outcomes of migrant workers. Their paper estimates a dynamic discrete choice model that incorporates stochastic job offers and job terminations. However, these studies do not consider how social networks affect search frictions. Gemici (2011) compares migration behaviors be-

[6] Pooling the assets to insure against the risk of the large agricultural income shocks across rural households is not common in China. On the one hand, since 2004, the government has set the minimum price for rice and wheat to minimize the aggregate rural income shocks. In rural China, each household has an equalized land size based on the household size, which causes no significant agricultural income differentials across households at the village level. On the other hand, the uncertainty of job search in urban areas in China is very low (e.g., the average duration of job search in the paper is about two months). These unique patterns make the activity of risk-sharing across households to insure against the risk of the large agricultural shocks is not common.

tween married couples and singles in a dynamic model of household migration with bargaining between family members. In her model, there exists uncertainty in the labor market, and search frictions influence individuals' migration decisions.

Colussi (2006) also studies the role of social networks on migration behavior through search frictions. He develops and estimates an equilibrium model to examine the impact of migrants' social networks on illegal Mexican immigration, allowing for repeated circular migration. He assumes migrants' networks can increase the probability of finding a job in the U.S. In his model, however, social networks cannot affect migration costs. Moreover, since social networks are defined as the number of Mexicans from their home village, he cannot consider individual investment in the social network.

This paper is also related to the literature analyzing why social networks are correlated with labor market outcomes. Munshi (2003) follows the idea of Carrington, Detragiache, and Vishwanath (1996) 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 an instrument. He finds that when the size of the network is larger, individuals are more likely to be employed and to hold a higher paying non-agricultural job.

In the theoretical literature, 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 networks. Chandrasekhar, Morten, and Peter (2020) find that the effect of the referral value on hiring decisions is nonmonotonic.

Besides the mechanism of social networks on migration decisions and the individuals' labor market outcomes, in this paper, I also allow the return and repeat migration behaviors in the model. The literature on migration studies of both "out-migration" and "return migration" is also known as circular migration. The seminal work of Kennan and Walker (2011) allows that individuals migrate temporary and also circular migration in the model, and it also allows for multiple location choices. Massey and Espinosa (1997) also shows that undocumented Mexican migrants tend to engage in repeated circular migration, moving back and forth between Mexico and the United States several times over their lifecycle. Thom (2010) examines the circular Mexican migration, and he structurally estimates the effects of changes in U.S. border enforcement and real exchange rates on Mexican migration. Most empirical studies of rural-urban migration in China do not consider the temporary migration behaviors (e.g., Zhao, 2003 and Zhang and Zhao, 2015). However, most rural individuals in China engage in temporary migration. For example, my sample from the Chinese Household Income Project data (2007-2009) shows that more than 40% of rural individuals who migrated to urban areas have experienced return or repeated circular migration.

3 Background, Data and Key Sample Statistics

3.1 Background of Rural-Urban Migration in China

Since 1958, the Chinese central government has restricted the mobility of the population through the household registration system (*hukou*). Given the birth location and the mother's household registration type, the child's *hukou* can be rural or urban, two kinds of household registration.[7] From 1958 to 1978, the rural people only who had job offers in urban areas or recruitment letters from universities could migrate from rural to urban areas. Since 1978, the people's commune system was replaced by the household-responsibility system, which loosened the restriction of rural residents' mobility. Although the central government had strict restrictions to limit working opportunities in cities, from 1979 to 1983, some rural residents began to migrate to work outside of their counties.

Between 1984 and 1988, the central government did not restrict rural-urban migration. At that time, under the planned economy, individuals needed to use food stamps to exchange food. However, if rural individuals migrated, they had to provide food stamps for themselves. Therefore, 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 was suspended between 1989 and 1991. After 1992, the government began to encourage rural-urban migration. In 2000, the government started to reform the household registration system to encourage more rural individuals to migrate.[8] For example, in 2007, 12 provinces in China canceled the rural household registration, which meant that rural individuals had the same household registration status as urban households in these provinces.[9] In these provinces, the local government does not distinguish between rural and urban residents any longer.

The easing of government restrictions on migration appears to have significantly affected people's migration decisions. In Appendix, Table A1 gives the inter-provincial migration in China from 1990 to 2005. There were 9.2 million people who migrated

[7]The rural and urban *hukou* are associated with different local benefits.

[8]A household registration officially provides an identification of a person as a resident of a specific area. The identification is based on information such as parents' or spouse's *hukou* status and birth location.

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

inter-province between 1990 and 1995. This number increased to 32 million between 1995 and 2000 and 38 million between 2000 and 2005. Figure 1 gives the approximate number of rural migrants since 2000.[10] The number of rural migrants increases from 78 million to 145 million within ten years.[11]

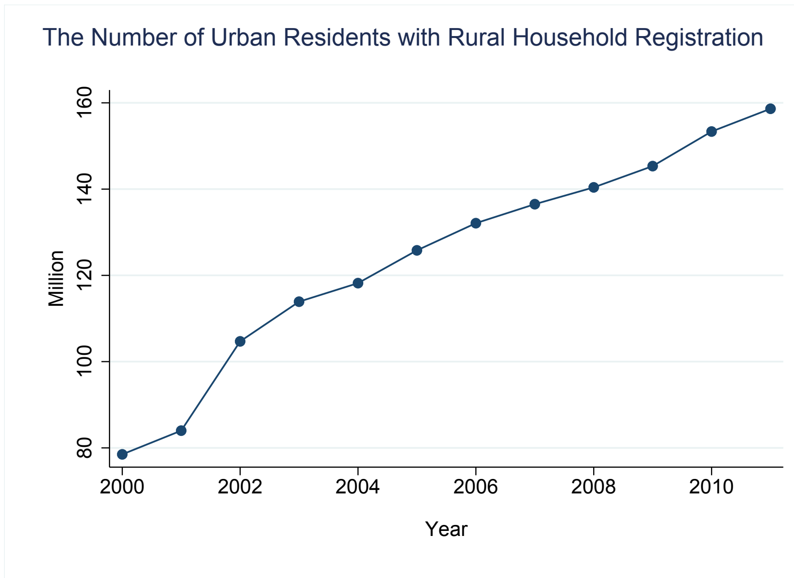


Figure 1: The Number of Stock Rural Migrants in China

After 2000, the central and local governments in China also proposed some policies to improve the working and living conditions of rural migrants. For example, in early 2000, several provinces and cities such as Guangdong, Beijing, Shanghai, and Xiamen started to set up social security schemes to cover rural migrants. In addition, a document issued by the State Council in May 2001 stated that local governments should provide nine years of compulsory education to migrant children through the public school system. However, until the end of 2006, only a few local governments have implemented this policy (Liang, 2007). Although the central and local governments in China tried to change the rural household system and the associated discrimination, Chan (2012) states that the effects of those policies have not been extensive.

The government's migration policy may affect individuals' location choices. In the CHIP data, individuals in different cohorts show different migration patterns.

[10] China Yearbook Rural Household Survey.

[11] All numbers referring to the measure of the migrants' number is stock value in this paragraph.

(a) (b)

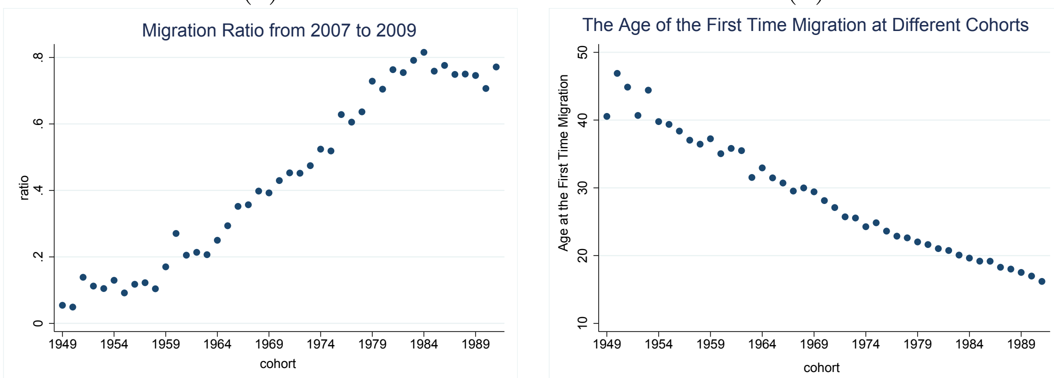


Figure 2: Internal Migration Patterns by cohorts

The definition of migration that I have in this paper is whether the urban residence location is out of his rural *hukou* (household registration) county. Figure 2(a) shows that the fraction of individuals who migrate to urban areas from 2007 to 2009 linearly increases across different cohorts. Figure 2(b) examines the average ages at first migration across different cohorts. It shows a clear pattern that average age at first migration decreases linearly with cohorts. [12]

Next, since the government policies change over time, I examine whether a year effect is also an important factor. The survival analysis is used to see whether both cohort and the year effects correlate with the average age of an individual's first-time migration. Table A2 gives the estimates, assuming a log-logistic distribution. The coefficient for education shows that individuals with a higher level of education migrate earlier. The year dummies are the time when the central government made a significant migration policy change. The year dummies (1984-1991) incorporate the policies for allowing migration, but still, people needed to use food stamps to exchange food. The year dummies (1992-2000) incorporate the transition period between planned and market economies. These two-year dummies are not significant at the 5% level. The year dummy (2001-2009) tries to capture the net effect of the policies made after 2000. The year dummy (2001-2009) is significant. Compared with cohort effects, year effects do not have a strong impact on individuals' migration choices.

[12] Both Figures 2(a) and 2(b) show that individuals in different cohorts have different migration patterns. The structural model introduces the cohort effects in the migration cost function to incorporate the government policies' differential effects across cohorts.

3.2 Data

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

CHIP's rural sample includes 8,000 rural households in 9 provinces. The nine provinces in the survey cover most provinces of migration origin and destination in China. In Appendix, Figure A2 gives a map of the nine provinces, including the net migration flow between 2000 and 2005. Table A1 shows that, from 2000 to 2005, more than 68% of migrants moved into these nine provinces, while 52% of migrants moved out of these nine provinces (NBS 2002, 2007).

In the analysis, I use the CHIP rural household survey sample. Individuals in the rural sample are all born in rural areas with rural household registration. The benefit of using the CHIP rural household survey is that the data follows all individuals regardless of their locations.[13] The panel is balanced, except for social network investment choices. Only the first two years' data contain information about social network investment.

In my analysis, I focus only on males to avoid issues with joint labor supply and fertility decisions. The sample contains information on work experience, job search duration, work locations, earnings, the presence of social networks, and social network investment decisions. Using this data, I construct the location choices, job search durations, and work statuses for the individuals between 16 and 60 years old for the three-year periods. Also, using the information of ever migration and the time of first migration can construct the history of location choices.

Using the rural sample data, I build the entire history of the work experience, no matter where the individuals are located. The data has the information about members in the household provided by the sampled individuals or their family members. For example, they provide when migrants left their home, when they return, and whether the destination is an urban county. Therefore, the location information allows me to construct the monthly location history for each individual in the rural sample.[14]

[13]The total number of men for the 8,000 households is 33,396, with 16,583 males. After restricting age between 16 in 2007 and 60 in 2009, the sample size shrinks to 11,385. In the data, 1,030 observations are missing information on fertility decisions or marital status. 1,099 observations are missing their work experience information from 2007 to 2009. The sample used for estimation includes 9,256 males in 6,400 households.

[14]In Section 4.8.2, I provide the details of how to deal with the case when individuals do not report the location information.

Table 1

Summary Statistics of the Estimation Sample		
	Sample	Observations
Fraction with networks	72.24%	9256
Network invest in 2007	61.26%	8026
	(0.49)	
Network invest in 2008	77.76%	7564
	(0.42)	
Education year	8.28	9256
	(2.14)	
Education level		9256
primary school or less	18.34%	
middle school	63.49%	
high school	14.55%	
college or above	3.62%	
Urban earnings	1220.23	
	(1168.82)	
Rural earnings	153.09	
	(477.38)	

1. Earnings have been adjusted by the location price index and the CPI price index.
2. Earnings are in Chinese currency yuan which is close to 1/8 of 1 U.S. Dollar in 2000.
3. Numbers in parentheses are standard deviations.

Table 1 displays selected descriptive statistics of the sample used in estimation. I restrict the sample to men who finished their formal education. The average years of schooling are 8.3 years. Since 1989, the central government has implemented the policy of nine years' mandatory education in China, which is equivalent to completing middle school (Grade 7 to Grade 9). About 18% of individuals have less than six years of schooling. Most people (63%) complete middle school. Only 4% of individuals have post-secondary education.

Social networks are defined as the presence of friends or relatives who are living in urban areas.[15] I define that social network investment as the activities of whether they send monetary gifts to their friends or relatives. In the survey, people answer

[15]In the CHIP data, social networks are measured at the household level. Individuals in the same household share the same social network.

whether they give gifts to their friends or relatives and the monetary values of gifts when applicable. In the data, the gifts also include those given to friends or relatives 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 possibilities introduce measurement errors in the social network investment. The estimation section provides details about how I deal with this measurement error problem.

As shown in Table 1, the variable of social networks is the presence of friends or relatives who are living in urban areas. More than 30% of individuals live in households that do not have social networks in urban areas. More than 60% of individuals invested in their social networks in 2007, and around 77% invested in 2008. I use the method proposed by Brandt and Holz (2006) to adjust earnings by location price index and the CPI price index. The base year is 2000. The average monthly earnings are around 1200 yuan, and the average earnings in rural areas are less than 200 yuan.[16] The earnings in urban areas are six times the earnings in rural areas.

Table 2 displays descriptive statistics for migrants and non-migrants. First, migrants have higher education levels in general than non-migrants. The education levels are higher for individuals with networks than those without networks among migrant and non-migrant groups. Second, migrants are much younger than non-migrants. The average age of migrants is 31. The average age of non-migrants is 42. The individuals with social networks are older than those without networks. Third, 60% of migrants get married, whereas 85% of non-migrants are married. Non-married individuals are more likely to migrate. Fourth, migrants with social networks have higher earnings than those without social networks. At the same time, migrants with networks have a smaller variance of earnings. Non-migrants' earnings do not show significant differences between those with and without social networks. Finally, the average job search duration is two months. Migrants without networks have a slightly longer job search duration than those with networks.

4 Model

In this section, first, I discuss the potential role of social networks in the labor market and develop the tests based on the empirical data to show why it is necessary to examine the quantitative effects based on the structural model. Second, given

[16]Yuan is the unit of the Chinese currency. The exchange rate between Yuan and the U.S. dollar was 8.28 in 2000.

Table 2: Descriptive Statistics for Migrants and Non-migrants

	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	1220.23	1220.23	
	(1168.82)	(1168.82)	
With Networks	1240.04	1240.04	
	(1232.21)	(1232.21)	
Without Networks	1162.12	1162.12	
	(909.06)	(909.06)	
Rural Earnings	153.09		153.09
	(477.38)		(477.38)
With Networks	158.81		158.81
	(480.50)		(480.50)
Without Networks	139.91		139.91
	(468.48)		(468.48)
Job Search Duration	2.20	2.20	
	(3.45)	(3.45)	
With Networks	2.19	2.19	
	(4.26)	(4.26)	
Without Networks	2.20	2.20	
	(2.99)	(2.99)	

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

the mechanism discussed, I construct a structural model to formally examine the impacts of social networks on individuals' outcomes.

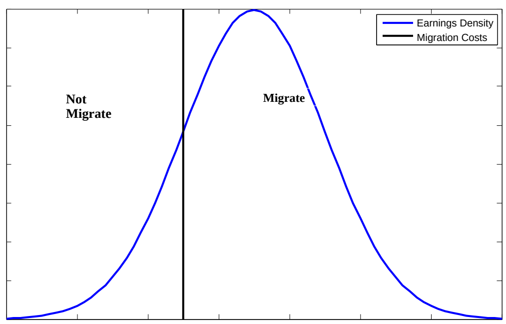
4.1 Mechanism of Social Network and Test

In Appendix B, I examine several correlations which are related to the key mechanisms proposed in this paper. I document a strong correlation between social networks and both migration choices and subsequent labor outcomes (see Tables B1-B2). In Table B3, I also show that social network investment is statistically significantly correlated with the presence of social networks.

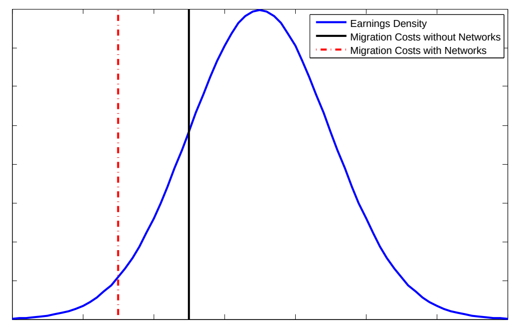
Before presenting the model, I illustrate why it is important to consider the impact of social networks through both migration costs and search frictions in a simple example. Figure 3(a) gives the expected urban earnings' distribution for rural individuals. The solid black line is migration cost net rural earnings. The individuals whose urban earnings are larger than the migration cost would migrate. Figure 3(b) illustrates how the problem would look when some individuals have social networks, and social networks only affect migration costs. In this case, the migrants with networks would have lower migration costs and have lower reservation earnings. Since the migrants with networks have lower migration costs, they are willing to migrate for lower earnings than those without networks. Therefore, we would observe that migrants with social networks have lower earnings than migrants without social networks.

Figures 3(c,d) illustrate how this problem would look when social networks only affect the job arrival rate. Figure 3(c) shows that individuals with social networks have a right-shifted expected earnings' distribution. In this case social networks only affect the job arrival rate, so individuals have the same migration cost. Figure 3(d) shows that migrants with social networks have higher earnings compared to ones without social networks. The reason is that higher job arrival rates increase the reservation values of taking job offers.

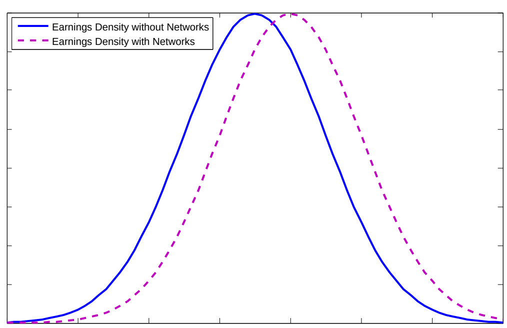
Given this intuition, I conduct the empirical test to examine which mechanism of social networks is consistent with the data (see Appendix C). For example, in Table C1, the testing results show that we should consider both of the two channels. The first two columns show that the coefficients of social networks on wages are statistically significantly positive when the selection term is not included. This result is consistent with the prediction that social networks only affect job arrival rates. However, when I control for the selection term, the coefficients of social networks (i.e., the columns under with correction) are consistent with the case that social networks only affect migration costs. These findings motivate that we should consider both



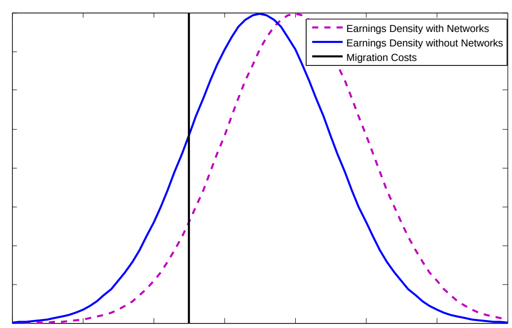
(a) With the Friction: Migration Costs



(b) Social Networks Affect Migration Costs



(c) With the Friction: Job Arrival Rate



(d) Social Networks Affect Job Arrival Rate

Figure 3: Illustration of How Social Networks Affect Migrants' Earnings

channels and quantify the impacts through different mechanisms.

Next, I present a model that allows me to quantify the importance of social networks to identify the mechanisms through which they operate and illustrate how individuals' investment decisions shape the patterns of social networks and migration patterns.

4.2 Model Specification

In this section, I model individuals' migration decisions, labor market transitions, and social network investment in a finite-horizon lifecycle framework. I account for unobserved (to econometrician) heterogeneity. To account for uncertainty regarding job offers, I model the migration decision problem is incorporated into a dynamic discrete time search framework. The decision period in this paper is taken to be one month in length.

In each period, an individual receives a location-specific flow utility from their earnings, unemployment benefits, psychic values of living in their hometown, and a choice-specific shock. Moving costs incur if they decide to change their locations and social network investment costs if they choose to invest in their social networks. The lifetime utility is given by the current utility flow and the discounted stream of expected future utilities. Uncertainty comes from migration costs, search frictions in the urban labor market, earnings transition, social networks' evolution, and idiosyncratic shock to utility.

4.3 Timing

Individuals' choices are made sequentially and also based on their current locations. The timing of individuals' decisions in the model is presented in Figure 4 and is as follows:

1. Individuals draw the marital and fertility shocks; marital and fertility transitions are exogenous and annual.

2. At the beginning of period t, the shock for social networks is realized, and individuals observe their social networks for the period.

3. (a) If individuals are unemployed in urban areas, job arrival (or immediate job offer) shocks are realized;

   (b) If individuals are employed in urban areas, separation shocks are realized.

4. (a) If individuals are in rural areas, both rural earnings' and migration cost shocks are drawn.
   (b) If individuals are in urban areas, both urban earnings' shocks and unemployment benefit shocks are drawn.
5. Following all of these shocks, location, employment, and network investment choices are made jointly.

4.3.1 Timing based on location

Individuals make decisions depending on their locations. Let inv_{it} be an indicator that takes value 1 if an individual invests in his social network, and 0 otherwise. Let $W_{it}^{j}(\text{inv}_{it})$ be the value of state $j \in \{e, r, n\}$ (e: employment in the urban area, r: rural, n: unemployment in the urban area). Denote V_{it}^{j} as follows:

$$V_{it}^{j} = \max\{W_{it}^{j}(1), W_{it}^{j}(0)\} \tag{1}$$

The decision process is described for each location separately:

Urban

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

1. If he just arrives in an urban area, he may receive an immediate offer. The individual chooses whether to invest in his social network simultaneously,
   (a) If he receives a job offer (with probability λ_{it}^{p}), he will choose between two options, either unemployment in the urban area, or accept the job offer ($\max\{V_{it}^{n}, V_{it}^{e}\}$).
   (b) If he does not get a job offer (with probability $1 - \lambda_{it}^{p}$), he will be unemployed in the urban area (V_{it}^{n}).[17]
2. If he has already stayed for one or more periods in the urban area, he gets a job offer with probability λ_{it}, which is affected by the presence of social networks $sn_{i,t}$. The individual chooses whether to invest in his social network simultaneously,

[17] In the data, there are above 30% of rural migrants whose job search duration is less than two weeks. Since the model period is one month, I introduce the immediate offer to match the job search duration in the data.

(a) If he gets a job offer, he will choose between three options: (1) unemployment in the urban area, (2) accept the job offer, or (3) return migration ($\max\{V_{it}^n, V_{it}^e, V_{it}^r - RM_{it}\}$). RM_{it} represents the return migration cost. The social network investment choice is made at the same time.

(b) If he does not get a job offer (with probability $1 - \lambda_{it}$), he will select between two options: (1) unemployment in the urban area or (2) return migration ($\max\{V_{it}^n, V_{it}^r - RM_{it}\}$). The social network investment choice is made at the same time.

3. If the individual works in an urban area, the exogenous job separation shock may hit him. The individual has to jointly choose whether to invest in his social network based on Equation (1).

(a) If he exogenously separates with the current job (with probability δ_i), he will choose to search for a job in an urban area or return migrate ($\max\{V_{it}^n, V_{it}^r - RM_{it}\}$).

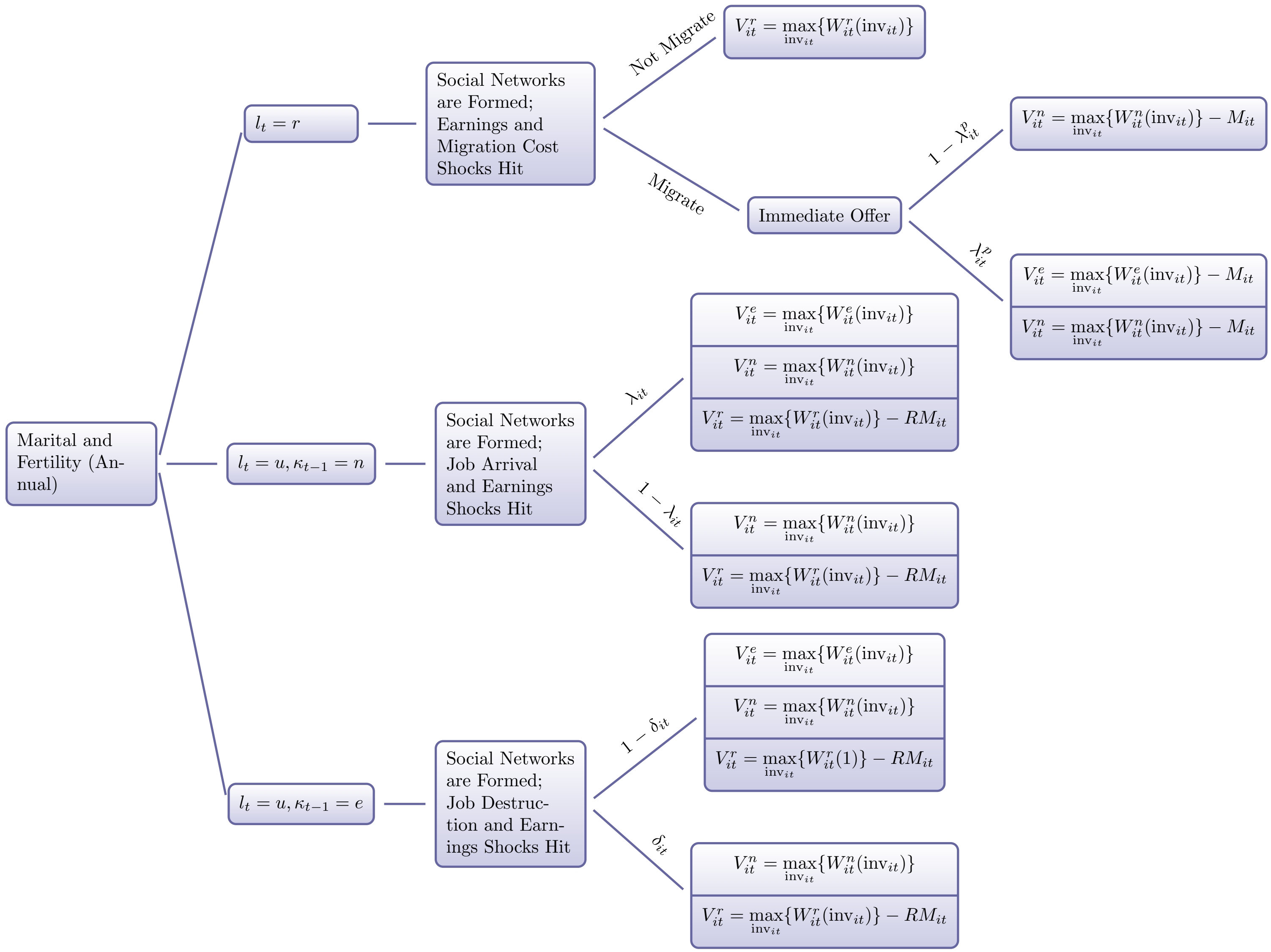
(b) If he does not exogenously separate with the current job (with probability $1 - \delta_i$), he will choose between three options: (1) keeping the current job, (2) quitting this job to unemployment in the urban area, or (3) quitting this job and returning home.

Rural

If he is in a rural area, his earnings ω_{it}^r are drawn from the distribution $G(\omega^r)$. Hence, he knows the value of living in the rural area (V_{it}^r) and the value of migration ($E\{\lambda_{it}^p \max\{V_{it}^e, V_{it}^n\} + (1 - \lambda_{it}^p)V_{it}^n\} - M_{it}$). Then, the migration decision is made based on equation 2:

$$\text{Mig}_{it} = \begin{cases} 1 & \text{if} \quad E\{\lambda_{it}^p \max\{V_{it}^e, V_{it}^n\} + (1 - \lambda_{it}^p)V_{it}^n\} - M_{it} > V_{it}^r \\ 0 & \text{else} \end{cases} \tag{2}$$

Where λ_{it}^p is the probability of getting an immediate offer and M_{it} is the migration cost. Social network investment choice is made jointly, based on Equation (1).



l_t is the location at the beginning of period t. κ_{t-1} is employment status at period $t-1$. $W^j(\text{inv}_{it})$ $j \in \{e, n, r\}$ is the choice specific value given the network investment decision.

Figure 4: The Timing of the Model

4.4 Model Specification

4.4.1 Earnings

Earnings are functions of education, work experience in rural and urban areas, endowment type, which is unobserved by econometrician, and location-specific idiosyncratic shocks. The log earnings of individual i with type k endowment, and in location $j \in \{u, r\}$ (u:urban, r:rural) at time t are described as

$$\omega_{it}^{jk} = \beta_0^{jk} + \beta_1^j \mathrm{S}_i + \beta_2^j \mathrm{exp}_{it}^r + \beta_3^j \mathrm{exp}_{it}^u + \beta_4^j {\mathrm{exp}_{it}^r}^2 + \beta_5^j {\mathrm{exp}_{it}^u}^2 + \varepsilon_{it}^j \tag{3}$$

In this paper the years of education S_i is predetermined, but individuals can accumulate human capital through learning by doing via location-specific work experience (i.e., rural exp^r, urban exp^u). Work experience in rural and urban areas are dependent on the history of endogenous decisions $\{d_{ik}\}_{k=1}^{t-1}$. The decisions include location, employment, and network investment choices. β_0^{jk} is type k individual's earning endowment at location j, $k \in \{1, \cdots, K\}$. Shock terms ε_{it}^j are assumed i.i.d. across individuals, locations, and time, and they are normally distributed with mean zero and variance σ_j^2, $j \in \{u, r\}$ (u:urban, r:rural). Individuals know the current period transient components (i.e., ε_{it}^j), but they do not know values of future transient components.[18] However, they do know the distribution of these future shocks and use them when taking expectations (i.e., rational expectations).

As specified in Equation (3), conditional on the unobserved endowment type, social networks do not directly affect earnings (i.e., the social network status does not enter the earnings equation). However, social networks may affect earnings indirectly through their effect on reservation earnings. On the one hand, individuals with social networks may have a higher job arrival rate, which will increase their reservation values for accepting urban job offers. If so, individuals with social networks will have higher accepted earnings. On the other hand, social networks may reduce migration costs. Hence, individuals with social networks are more likely to migrate for the same expected earnings because of lower migration costs. This implies that individuals with social networks may have lower reservation values of taking urban job offers than those without social networks. From the discussion above, the net relationship between social networks and migrants' earnings is not clear since the effect of social networks goes through different channels (i.e., lower migration costs and higher job arrival rates).

[18] Alternatively, one would to extract individual components of uncertainty as in Navarro and Zhou (2017).

4.4.2 Social Networks

In the model, individuals can invest in their social networks to strengthen connections with their friends (e.g., they may give gifts to their friends or contact them by phone or mail). For an individual i with type k, social networks are formed according to the following dynamic probit model:

$$sn_{it}^k = \begin{cases} 1 & \text{if} \quad \beta_0^{sk} + \beta_1^s \text{inv}_{it-1} + \beta_2^s \text{mar}_{it} + \beta_3^s \text{child}_{it} + \beta_4^s sn_{it-1}^k + \varepsilon_{it}^s > 0 \\ 0 & \text{else} \end{cases} \tag{4}$$

where sn_{it}^k is the indicator of social networks status at period t. It takes value 1 if the individual has social networks and 0 otherwise. inv_{it-1} is the individual's social network investment decision at period $t-1$. $\text{inv}_{it-1} = 1$ means that he invests in his social networks at period $t-1$, otherwise $\text{inv}_{it-1} = 0$. If the coefficient of β_1^S is positive, he increases the probability of having social networks by the investment choice (e.g., giving gifts). mar_{it} is marital status at period t and child_{it} is the number of children at period t. β_0^{sk} is social network endowment if individual is type k, $k \in \{1, \cdots, K\}$. The shock ε_{it}^s is i.i.d. across individuals and time and with standard normal distribution. Individuals cannot observe future shock terms ε_{it}^s, but they know the distribution of shocks.

In this model, the investment decision is a discrete choice that depends on the trade-off between the gain from increasing the probability of having social networks and the cost of investing. The key mechanism of investing in social networks is that it may increase the probability of having (or keeping) social networks to reduce migration costs and increase the job arrival rate. Individuals do not know their future shocks, so their investment choices are based on their expectations of future shocks.

4.4.3 Migration and Return Migration Costs

If individuals migrate between rural and urban areas, they have to pay migration costs. One of the proposed channels through which social networks operate in the model is that they may affect migration costs directly.[19] Migration costs M_{it} depend on the current period's presence of social networks, marital status, number of children, birth cohort, endowment value from unobserved type, and migration cost shock. I assume asymmetric migration costs: migration costs (M_{it}) may not equal to

[19] Carrington, Detragiache, and Vishwanath (1996) build a dynamic macro model to examine the role of social networks on migration decisions. They also assume social networks reduce migration costs.

return migration costs (RM_{it})[20]. Migration and return migration costs in this paper are specified by the Equations (5)-(6):

$$\mathrm{M}_{it}^{k} = \beta_0^{mk} + \beta_1^{m} sn_{it} + \beta_2^{m}\mathrm{mar}_{it} + \beta_3^{m}\mathrm{child}_{it} + \beta_4^{m}\mathrm{cohort}_i + \varepsilon_{it}^{m} \tag{5}$$

$$\mathrm{RM}_{it}^{k} = \beta_0^{rmk} + \beta_1^{rm}\mathrm{mar}_{it} + \beta_2^{rm}\mathrm{child}_{it} + \beta_3^{rm}\mathrm{cohort}_i \tag{6}$$

where sn_{it} is the indicator of the presence of social networks in urban areas at period t.

In the model, I also allow different cohorts to have different migration costs to accommodate the fact that rural individuals across different cohorts have different migration patterns. Table A2 shows that younger cohorts are more likely to migrate earlier than older cohorts. The cohort term attempts to capture the net effect of the change of migration policies over four decades in China.[21]

4.4.4 Job Arrival and Destruction Rates

When people migrate to urban areas, they have to search for jobs from the unemployed state. Social networks may help individuals reduce search frictions in urban areas. The probability of getting an immediate offer λ_{it}^{p} upon arrival to the urban area, and the job arrival rates λ_{it} in urban areas are parameterized as:

$$\lambda_{it}^{pk} = \frac{\exp\{\beta_0^{lpk} + \beta_1^{lp} 1_{sn_{it}} + \beta_2^{lp}\mathrm{S}_i\}}{1 + \exp\{\beta_0^{lpk} + \beta_1^{lp} 1_{sn_{it}} + \beta_2^{lp}\mathrm{S}_i\}} \tag{7}$$

$$\lambda_{it}^{k} = \frac{\exp\{\beta_0^{lk} + \beta_1^{l} 1_{sn_{it}} + \beta_2^{l}\mathrm{S}_i\}}{1 + \exp\{\beta_0^{l} + \beta_1^{l} 1_{sn_{it}} + \beta_2^{l}\mathrm{S}_i\}}. \tag{8}$$

To model exogenous job separation, the job destruction rate is parametrized as:

$$\delta_{it}^{k} = \frac{\exp\{\beta_0^{\delta k} + \beta_1^{\delta}\mathrm{S}_i\}}{1 + \exp\{\beta_0^{\delta k} + \beta_1^{\delta}\mathrm{S}_i\}}. \tag{9}$$

[20] To reduce the computation burden, I assume that return migration costs follow the given parameterized equation.

[21] I model the cohort effect as linear. Figure 2(b) shows the average age of first migration across different cohorts. There is a linear relationship between cohort and the average age of first migration.

4.4.5 Marriage and Fertility Transition Process

I model annual marriage and fertility transitions with an exogenous process. The marital transition process follows as a continuous duration model with log-logistic distribution.[22] The survival function is:

$$\text{Sur}_i(t) = (1 + (e^{-(\beta_0^{ma} + \beta_1^{ma} S_i)} t)^{1/\gamma})^{-1}. \tag{10}$$

Here, $\text{Sur}_i(t)$ is the probability of being single at age t, S_i is years of education, and γ is the parameter of the loglogistic distribution. Then, the conditional probability of getting married at period t is given by:

$$\Pr(\text{mar}_{it} = 1 | \text{mar}_{it-1} = 0) = \frac{\text{Sur}_i(t) - \text{Sur}_i(t+1)}{\text{Sur}_i(t)}$$

Fertility is determined by the following equation:

$$F_{it} = \begin{cases} 1 & \text{if} \quad \beta_0^f + \beta_1^f \text{age}_{it} + \beta_2^f \text{age}_{it}^2 + \beta_3^f \text{child}_{it} + \beta_4^f \text{child}_{it}^2 + \beta_5^f \text{S}_i + \beta_6^f \text{mar}_{it} + \varepsilon_{it}^f > 0 \\ 0 & \text{else} \end{cases}. \tag{11}$$

Equation (11) shows that fertility is correlated with age, the number of children, years of education, and marital status.

4.4.6 The Flow Value of Living in Rural Areas

The per-period utility in rural areas for individual i, at time t is given by

$$u_{it}^{rk} = \omega_{it}^{rk} - \nu \text{inv}_{it} + \phi_{it}^k, \tag{12}$$

where ω_{it}^{rk} is rural earnings for individual i at time t, ν is the cost if he invests in his social networks at time t and ϕ_{it}^k is the psychic value of living in the rural area, which is given by

$$\phi_{it}^k = \beta_0^{\phi k} + \beta_1^{\phi} \text{age}_{it} + \beta_2^{\phi} \text{age}_{it}^2 + \beta_3^{\phi} \text{mar}_{it} + \beta_4^{\phi} \text{child}_{it}. \tag{13}$$

I introduce a psychic value of living in rural areas is because, as documented in the migration literature (e.g., Kennan and Walker, 2011), people seem to place an added value to their home towns, especially older individuals. $\beta_0^{\phi k}$ is the endowment value if the individual is type k, $k \in \{1, \cdots, K\}$.

[22] I assume there is no divorce for this marital transition. The annual divorce rate in rural areas is lower than 0.1% in the data.

4.4.7 The Flow Value of Living in Urban Areas

Unemployment State

The per-period utility of being unemployed in urban areas for individual i, at time t is given by

$$u_{it}^{nk} = \xi_{it}^{k} - \nu \mathrm{inv}_{it}, \tag{14}$$

The per-period utility of being unemployed in urban areas for individual i, at time t is given by where ξ_{it}^{k} is the per-period utility of being unemployed in urban areas, and ν is the cost of investing in social network investment. As shown in Equation (15), the per-period utility of being unemployed (ξ_{it}^{k}) is assumed to be a function of the individual's age, marital status, and the number of children, and individual's endowment. This setting allows older people to have difficulties assimilating to a new environment, so they may have different valuations of being unemployed in urban areas. Marital status and the number of children reflect the net value for an individual to live with his family.

$$\xi_{it}^{k} = \beta_0^{\xi k} + \beta_1^{\xi} \mathrm{age}_{it} + \beta_2^{\xi} \mathrm{age}_{it}^2 + \beta_3^{\xi} \mathrm{mar}_{it} + \beta_4^{\xi} \mathrm{child}_{it} + \varepsilon_{it}^{\xi}. \tag{15}$$

The shock term ε_{it}^{ξ} are assumed i.i.d. across individuals and time.

Employment State

The per-period utility of being employed in urban areas for individual i, at time t is given by

$$u_{it}^{ek} = \omega_{it}^{uk} - \nu \mathrm{inv}_{it}, \tag{16}$$

where ω_{it}^{uk} is urban earnings for individual i at time t.

4.4.8 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 include age, years of education, marital status, number of children, accumulated work experience in rural and urban areas, the presence of social network, social network investment at period $t-1$, and unobserved type. 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 Markovian and denote its transition probability by $\Pr(H_{it+1}|H_{it}, D_{it})$.

The process of social networks status is given by a dynamic probit model. Work experience in rural and urban areas is determined by the action history $D_{it} = \{\{d_{it}^m\}_{t=1}^T\}$, where $\{d_{it}^m\}_{t=1}^T$. $m \in \{1, \cdots, 5\}$ (1: migrate, 2: employed in urban, 3: employed in rural, 4: unemployed in urban, and 5: return migrate).[23]

4.5 Value Function

The choice-specific Bellman equations for each of the three states are

$$\begin{aligned}
W_{it}^n(\mathrm{inv}_{it}; H_{it}) &= u_{it}^n + \frac{\lambda_{it}}{1+\rho} E(\max\{V_{it+1}^n(H_{it+1}), V_{it+1}^e(H_{it+1}), V_{it+1}^r(H_{it+1}) \\
&\quad -RM_{it+1}\}|H_{it}) + \frac{1-\lambda_{it}}{1+\rho} E(\max\{V_{it+1}^n(H_{it+1}), V_{it+1}^r(H_{it+1}) \\
&\quad -RM_{it+1}\}|H_{it}) \\
W_{it}^e(\mathrm{inv}_{it}; H_{it}) &= u_{it}^e + \frac{\delta_{it}}{1+\rho} E(\max\{V_{it+1}^n(H_{it+1}), V_{it+1}^r(H_{it+1}) - RM_{it+1}\}|H_{it}) \\
&\quad + \frac{1-\delta_{it}}{1+\rho} E(\max\{V_{it+1}^e(H_{it+1}), V_{it+1}^r(H_{it+1}) - RM_{it+1}, V_{it+1}^n(H_{it+1})\}|H_{it}) \\
W_{it}^r(\mathrm{inv}_{it}; H_{it}) &= u_{it}^r + \frac{1}{1+\rho} E(\max\{V_{it+1}^r(H_{it+1}), \lambda_{it+1}^p \max\{V_{it+1}^n(H_{it+1}), V_{it+1}^e(H_{it+1})\} \\
&\quad +(1-\lambda_{it+1}^p) V_{it+1}^n(H_{it+1}) - M_{it+1}(H_{it+1})\}|H_{it}).
\end{aligned}$$

where $W_{it}^j(\mathrm{inv}_{it}; H_{it})$ is the value of state $j \in \{e, r, n\}$ given the social network investment decision (e: employment in urban areas, r: rural, n: unemployment in urban areas). Based on the choice specific Bellman equation, the value function is stated by the following equation:

$$V_{it} = \begin{cases} \lambda_{it} \max\{V_{it}^n, V_{it}^e, EV_{it}^r - RM_{it}\} + (1-\lambda_{it}) \max\{V_{it}^n, EV_{it}^r - RM_{it}\} & \text{if unemployed in urban} \\ (1-\delta_{it}) \max\{V_{it}^n, V_{it}^e, EV_{it}^r - RM_{it}\} + \delta_{it} \max\{V_{it}^n, EV_{it}^r - RM_{it}\} & \text{if employed in urban} \\ \max\{V_{it}^r, E\{\lambda_{it}^p \max\{V_{it}^n, V_{it}^e\} + (1-\lambda_{it}^p) V_{it}^n\} - M_{it}\} & \text{if in rural} \end{cases} . \tag{17}$$

To simplify notations, I denote $V_{it}^j = \max\{W_{it}^j(1), W_{it}^j(0)\}$, $j \in \{e, r, n\}$. u_{it}^j, $j \in \{e, r, n\}$ is the flow utility at different states. M_{it} is migration cost, and RM_{it} is

[23] Social network investment choice is made jointly with these choices.

return migration cost. λ_{it}^{p} is the probability of taking an immediate offer. λ_{it} is the job arrival rate and δ_{it} is the job destruction rate.

4.6 Identification

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

The endowment under different unobserved types can be identified from the consistent behavior observed from panel data. For example, through earning equations, conditional on the same observables, if two individuals have persistent large earning gaps, it will identify endowment values under different unobserved types.

The job arrival rate λ_{it} can be identified using the information of the unemployment durations in urban areas. The probability of getting an immediate offer λ_{it}^{p} can be identified from the fraction accepting a job offer when they just arrive in urban areas. Since the fraction of employed migrants who switch into unemployment can be observed, the job separation rate δ_{it} can be identified from the behavior of leaving high salary urban jobs to the unemployment state or back to home locations.

Since I cannot identify both psychic values of living in rural and urban areas, I normalize the non-pecuniary value of living in urban areas to zero. Under this normalization, the psychic value of living in rural areas can be identified as follows. Since the data provide the information about the choices of locations are not completely explained by variation in earnings across locations. For example, older individuals are more likely to stay in rural areas. The observed variables are correlated with observed location choices after controlling for earnings; this provides information about the impact of the observed variables on the psychic value of living in rural areas ϕ_{it}.

Next, the psychic value ϕ_{it} of living in rural areas and migration costs M_{it} can be separately identified since an exclusive variable cohort in migration cost function but not in psychic value. Therefore, I can use the cohort variation to separately identify migration cost and the psychic value of living in rural areas.

About the identification between migration cost and search frictions (i.e., job arrival rate): since the job arrival rate λ_{it} can be identified from the migrant's non-first spell of job search. For example, suppose migrants separate from the firm in urban areas and become unemployed. In that case, the spell between unemployed and a new firm in an urban area could help to identify the job arrival rate λ_{it}. In this paper, I assume migration costs are monetary costs and do not include time costs. Based on this assumption, the migration cost M_{it} can be separately identified from the immediate job offer probability λ^p_{it}.

4.7 Objective Function and Solution Method

Individuals maximize their present discounted values of lifetime utility from the year they finish their education to a terminal age, $t = T$. Denote the utility associated with each choice as u^m_{it}. Then individuals make choices to maximize their objective function $V_{it}(H_{it})$:

$$V_{it}(H_{it}) = \max_{D^m_{it}} u^m_{it}(H_{it}) + \frac{1}{1+\rho} E(V_{it+1}(H_{it+1})|H_{it}). \tag{18}$$

The expectation operator E in equation (18) is taken with respect to the joint distribution of stochastic shocks $\varepsilon^{\xi}_{t+1}, \varepsilon^{u}_{t+1}, \varepsilon^{r}_{t+1}, \varepsilon^{m}_{t+1}$, ε^{s}_{t+1}, the probability of receiving a job offer, job separation, getting married, and having a child.

Given the finite horizon, the model is solved numerically through backward recursion of the Bellman equation. This procedure, however, cannot be applied directly due to the high dimensionality of the problem. Furthermore, the decision period in the model is a month which brings an additional computation burden. To reduce the computation burden associated with the high dimensionality of the problem, I adopt an approximation method similar to the one employed in Keane and Wolpin (1994). Instead of calculating continuation values at all points of the state space, I approximate them using a polynomial on the states. At each t, I calculate the Emax functions (i.e., the continuation values) for a subset of the state space and estimate a regression function as a polynomial in those state space elements. I use the predicted values from the regression to approximate the alternative-specific value functions given by Equation (18).

4.8 Estimation

4.8.1 Likelihood

The model is estimated by 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., $\{D_{it}^m : m = 1, \cdots, M\}$)
- Outcomes: earnings, presence of social networks

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 of choices (i.e., migration, employment, return migration, and network investment) and outcomes at each period t.

To allow that individuals have different levels of abilities, I follow the ideas of Heckman and Singer (1984) and Keane and Wolpin (1997) and assume that ability is unobserved by econometricians. However, econometricians know the total number of types (K). π_i^k is the probability that individual i belongs to type k, which is the function of an individual's years of education. [24]

Let $t_{0i} \in [2007, 2009]$ denote the first period an individual is observed in the data. Notice that the state space includes both lagged variables (inv_{it-1}) and accumulated rural and urban work experience (expr_{it}, expu_{it}). These variables, however, are not observed for all individuals in the data. Let $\bar{H}_{0i}$ denote the value of the state space when an individual enters the sample. Then, if the probability of $\bar{H}_{0i} = \bar{h}_{t_{0i}}$ were known, the contribution to the likelihood for individual i would be:

$$\Pr(c(t_{0i}), \cdots, c(t_{Ti})|\bar{H}_{it}) = \sum_{k=1}^{K} \pi_i^k \left[\sum_{\bar{H}_{0i} \in \Omega} \Pi_{t=t_{0i}}^{t_{Ti}} \Pr(c(t)|H_{it}, d_{it}) \Pr(\bar{H}_{t_{0i}}|type = k) \xi(p_m) \right] \tag{19}$$

Equation (19) assumes that we know $\Pr(\bar{H}_{t_{0i}})$ and hence we can integrate it out. This, however, is not the case. It further assumes that the measurement of social network investment in the data corresponds exactly to the one in the model, which is not the case either. In the next two sections, I describe a methodology to deal with these two problems.

4.8.2 Initial Condition Problem

As I discussed, to calculate the likelihood for each individual, I need the state variables in the year when he enters the sample. The data provides the whole marriage and fertility

[24] The probability is parameterized as follows: $\pi_i^k = \frac{exp(\beta_{\pi,1}^k edu_i + \beta_{\pi,2}^k)}{1+exp(\beta_{\pi,1}^k edu_i + \beta_{\pi,2}^k)}$

histories for each individual. However, the data does not cover the information on work experience in rural and urban areas for some individuals. I use a simulation method to solve this missing history problem. The basic idea is that, for the current value of the parameters, I simulate the transient shocks and use the value functions to simulate individuals' sequential decisions from an initial period until they enter the sample. I then simulate such histories R times to calculate the probability of observing an individual with $H_{t_{0i}} = h_{t_{0i}}$ to contribute the likelihood. The specific procedure is :

1. Given the current values of the parameters, the model is solved on grids, and the value functions are saved.
2. Given the value functions, I draw from the distribution of the shocks to simulate a history from the time when the individual finishes formal education to the time he enters the sample t_{0i} for individual i. Let the value of the state at t_{0i} implied by simulation $r = 1, \cdots, R$ be denoted by $\bar{H}^r_{t_{0i}}$.
3. Repeat step (2) R times.[25]
4. Given $\{H^r_{t_{0i}}\}^R_{r=1}$, calculate $\Pr(H_{t_{0i}})$, which is needed to be calculated in the likelihood Equation (19).

4.8.3 Measurement Errors

As described in the data section, there are likely measurement errors for the social network investment variable. Rural households only report whether they send gifts to their best five friends or relatives in the data. This leads to two types of measurement errors: they may send gifts to someone outside of the best five friends or relatives, and/or the gifts may be given to friends or relatives living in rural areas. In the estimation, I assume the probabilities of having each type of measurement error (p_m) are the same, and measurement errors are independent with all shocks and unobserved types. Since I only observe the investment choices in 2007 and 2008, the likelihood function for individual i is given by :

$$\begin{aligned}\Pr(c(t_{0i}), \cdots, c(t_{2009})|\bar{H}_{it}, type = k) \quad = \quad & \sum_{\bar{H}_{2007}\in\Omega} \Pr(\bar{H}_{2007})\Pi^{t_{2007.12}}_{t=t_{2007.1}} \Pr(c(t)|H_{it}, d_{it}, type = k) \\ & ((1-p_m)^{m_{07}=a_{07}} p_m^{m_{07}\neq a_{07}})\Pi^{t_{2008.12}}_{t=t_{2008.1}} \Pr(c(t)|H_{it}, d_{it}, type = k) \quad (20) \\ & ((1-p_m)^{m_{08}=a_{08}} p_m^{m_{08}\neq a_{08}})\Pi^{t_{2009.12}}_{t=t_{2009.1}} \Pr(c(t)|H_{it}, d_{it}, type = k),\end{aligned}$$

where $m_j, j \in \{2007, 2008\}$ is the model prediction of social network investment at

[25] I simulate 500 times for each individual who misses work experience information.

period j, and $a_j, j \in \{2007, 2008\}$ is the data measure of social network investment at period j.

4.8.4 Estimation Procedure

The estimation algorithm is developed to incorporate both the initial condition and measurement error problems, and the assumed exogenous stochastic processes for marital status and fertility. The procedure is as follows :

1. Estimate the exogenous marital and fertility stochastic process and get the parameters Θ_1.[26]
2. Given Θ_1, and the initial guess for the other parameters Θ_2, the model is solved on grids, and the value functions are approximated as described in section 4.7.
3. For the individuals missing the value of the state, I draw the shock terms and simulate their choices for the missing periods to calculate $\Pr(H_{t_{0i}})$ as described in section 4.8.2; consider the measurement errors as described in section 4.8.3.
4. Calculate the likelihood and update the parameters Θ_2.
5. Repeat from Step 2 to Step 5 until parameters Θ_2 converge.

5 Empirical Application

5.1 Estimation

5.1.1 Model Fit

In this section, I evaluate the model fit by using the estimates to simulate individuals' behaviours and compare the simulated results to the data moments. Notice that I use the method of maximum likelihood to estimate the model. Therefore, all moments reported in Tables D5 and D6 are not targeted in the estimation. Tables D5 and D6 give the comparison between the model predictions and data moments.[27]

[26] Table D1 gives the estimates of parameter Θ_1.

[27] For each individual, I simulate 50 times and the results reported are the mean of simulation results. I simulate the decisions for each individual from the age of finishing formal education to the age of 60. The moments calculated are based on the simulation results from the year 2007 to 2009.

Table D5 shows the comparison to the earnings' moments. The data column gives the selected data moments for both migrants' and non-migrants' earnings, including the mean and variance of log earnings. The other column gives the simulated moments based on the model estimates. Although the simulated standard deviations of log earnings are slightly larger than those in the data, the calculated moments fit the data quite well. The model can successfully capture that earnings with networks are higher than the earnings without networks.

Table D6 gives the model fit for choices. The simulated moments can fit the fraction of the individuals with networks quite well. For example, in the data, 72.2% of individuals have social networks. The simulated moments are 71.6% in the model.

When examining the composition of the migrants, we find that both the models can capture rural individuals' migration choices quite well. For example, in the data, 29.0% of rural individuals live in urban areas, and the model prediction is 29.3%. When examining the decomposition of rural individuals living in urban areas, we can find that the model also matches the moments well. Among the 29.0% of rural individuals who live in urban areas, 21.8% have social networks and 7.2% do not. The model predicts this decomposition as 21.2% and 8.0%. Conditional on migrants' employment status, 26.5% of the individuals work in urban areas and 2.5% of them are unemployed in urban areas. The model predictions are 27.0% and 2.2%, respectively.

From Table D6, we see that the job search duration matches quite well: in the data, the average job search duration is 2.20 months; the model predicts the job search duration is 1.83 months. The model also sufficiently captures the behaviours of accepting immediate offers after migrating. For example, there are 0.38% of rural individuals who get a job immediately after migrating in the data. The model predicts 0.32%. As the table shows, the model also fits the moments related to return and repeat migration reasonably well.

5.1.2 Estimates

The estimated parameters are reported in Tables D2 and D3 (see Appendix D). The parameter estimates are consistent with what one would expect. In particular, the signs of effects of social networks on these two channels are the same as we expect. First, the effect of social networks on migration costs is negative and statistically significant. This means that the presence of social networks reduces migration costs (e.g., they help migrants to settle down in urban areas). From the migration cost

Table 3: The Impact of Social Networks

	Key Estimates and Average Effects	
	Migration Cost	Job Arrival Rate
Social Networks Coefficient	-0.55^{***}	0.56^{***}
Average		
With Networks	528.61	0.45
Without Network	566.78	0.32

1. The unit of migration costs for rural migrants is 1 Yuan.
2. The first panel in the job arrival rate column presents the point estimate for social networks in the job arrival rate equation. The second panel shows the calculated average job arrival rate for rural migrants who are not employed depending on their network status.
3. *** stands for ($p < 0.001$).

equation, we find that married individuals and those with children have higher migration costs. Also, the older cohorts have higher migration costs. Second, social networks significantly increase the job arrival rate. The estimates are consistent with the idea that migrants can get job information from their friends or relatives in urban areas.

Table 3 displays the role of social networks through the two channels. The model estimates show that the average migration costs for individuals with networks are 93.3% of the value for those without networks. Social networks may also affect search frictions. From the estimates of the job arrival rates, we see that social networks can significantly reduce search frictions. The average arrival rate for the individuals with networks ($\bar{\lambda} = 0.45$) is 40% higher than that for those without ($\bar{\lambda} = 0.32$).

The estimates of the earning equations in Table D2 show that the large gap between urban and rural areas does not mainly come from the returns to years of education and work experience. The difference mainly comes from the constant term, which may be explained by the fact that the productivity in urban areas is higher than that in rural areas. The return of education for rural migrants in urban areas is even lower than the return in rural areas. In rural areas, the impact of an additional year of schooling on earnings (2.7%) is higher than the one in urban areas (2.3%), which is consistent with the finding in Navarro and Zhou (2020). The return of an additional month of urban work experience to urban earnings is higher than the return of an additional month of urban experience to rural earnings. Similarly, the return of an additional month of rural experience is higher to rural earnings than that to urban earnings.

From the unemployment equation, married individuals and those with more children have a higher value on unemployment than single or childless individuals. At the same time, the estimates show that older people have a higher flow of utility values if they are unemployed. The coefficients in the equation of the psychic value of living in rural areas show that older people have a higher flow utility of living in rural areas. This finding is consistent with the migration literature (e.g., Kennan and Walker, 2011) and the data observation that migrants are younger than those individuals living in rural areas.

From Tables D3 and D4, years of education is positively correlated with type 2. When an individual has a lower level of education, he is more likely to be type 1 or type 3. For example, in Table D4, when the year of education is less than 1, the probability of being type 2 is 20%; however, when the year of education is 16, the probability of being type 2 increases to 59%. Figure 5(a) shows that type 1 and 2 migrants' density of urban earnings is very close, and type 3 migrants have lower urban earnings. Figure 5(b) shows the order of rural earnings conditional on different unobserved types: type 2 individuals have the highest rural earnings, and type 3 has the lowest rural earnings.

Although type 3 individuals are more likely to be less productive, from Figures 6(a) and (b), we find that type 3 individuals have less migration costs and higher job arrival rates. In terms of the psychic values of living in rural areas, the density for the three types is quite similar, and the type 2 individuals have slightly lower psychic values of living in rural areas as we can see in Figure 6(c).

When comparing these three unobserved types, type 1 and 2 perform quite similarly when they are in urban areas. They have very close urban earning densities, and the density of migration costs and job arrival rates are also very close. The biggest difference between type 1 and type 2 is from the outcomes in rural areas. For example, type 2 individuals have higher rural earnings than type 1, and type 1 individuals have higher psychic values of living in rural areas than those for type 2.

(a)
Rural Migrants' Earnings

(b)
Rural Earnings

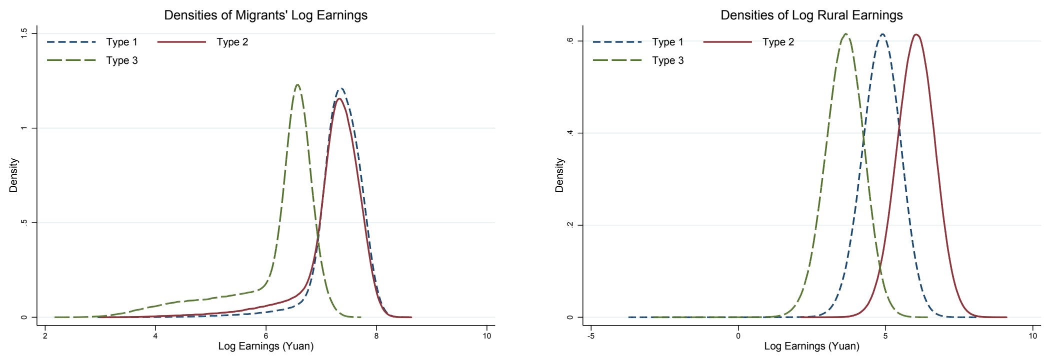


Figure 5: Distribution of Log-earnings by Unobserved Types

5.2 Decomposition Analysis

In this section, I conduct counterfactual simulations to decompose the effects of social networks through migration costs and the job arrival rate. I then examine how social networks affect rural individuals' migration and social network investment choices.

To assess the effects of social networks on migration costs and labor market search frictions, I simulate the model under three different restrictions on the parameters. In the first specification, I turn off the effects of social networks on both channels, (i.e., $\beta_1^m = 0$, $\beta^{l_1^p} = 0$, $\beta_1^l = 0$). In this case, social networks play no role in the model. In the second specification, I turn off the effects of social networks on the job arrival rate (i.e., $\beta^{l_1^p} = 0$, $\beta_1^l = 0$). That is, social networks are only allowed to affect migration costs. In the third specification, I turn off the effects of social networks on migration costs (i.e., $\beta_1^m = 0$).

(a) Migration Costs | (b) Job Arrival Rates | (c) Psychic Values

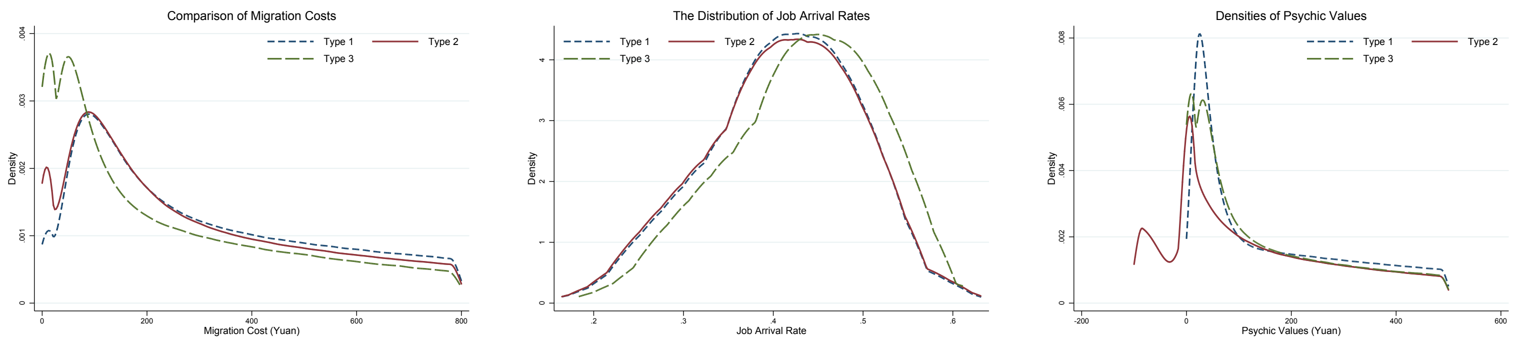


Figure 6: Distributions of Migration Costs, Job Arrival Rates, and Psychic Values by Unobserved Types

Table 4: Counterfactual Results: Social Networks

	Model	Social Networks Only Affect		
		Neither	Migration Cost	Job Arrival Rate
Fraction of Individuals with Networks	71.64%	68.73%	69.49%	70.08%
Fraction of Migrants	29.26%	15.39%	16.60%	27.04%
With Networks	21.20%	10.15%	11.59%	19.02%
Without Networks	8.06%	5.24%	5.02%	8.03%
Fraction of Non-migrants	70.74%	84.61%	83.39%	72.96%
With Networks	50.44%	58.57%	57.90%	51.06%
Without Networks	20.30%	26.04%	25.49%	21.89%
Fraction of Unemployed Migrants	2.22%	0.13%	1.38%	2.08%
Job Search Duration	1.83	1.80	1.78	1.84

1. Migrants include people who were born in rural areas and resided in urban cities who can be employed or unemployed.
2. The model column shows the benchmark values simulated by the model estimates.
3. The neither column gives the simulation results when social networks affect neither migration costs nor the job arrival rate.
4. The migration cost column gives the counterfactual results if social networks only reduce migration costs.
5. The job arrival rate column gives the counterfactual results if social networks only increase the job arrival rate.

Table 4 presents the decomposition simulation results for the model in terms of the role of networks (i.e., neither of two channels, only affect migration costs, or networks only affect the job arrival rate). The second column presents the unrestricted (i.e., allowing social networks to affect both channels) predictions to use as a baseline case. The third column shows the model prediction if social networks do not affect either the job arrival rate or migration costs. Without the effect of social networks, only 15% of rural individuals will live in urban areas. The difference is more than 13 percentage points comparing to the results where network effects exist (29%). When social networks only affect migration costs (Column 4), 17% of rural individuals will live in urban areas. The job arrival rate column shows that if social networks only increase the job arrival rate and have no effect on migration costs, 27% of rural individuals migrate. These simulation results show that social networks reduce search frictions is much larger than reducing migration costs.

To examine whether the social network affects earnings as the model stated, I run wage regression under different counterfactual analyses. Table 5, column 1 shows the correlation between the social network and rural migrants' earnings when social network only affects migration costs. Under this condition, individuals with social networks have lower reservation values, and we may expect a negative correlation

Table 5: Correlation between Social Networks and Log Earnings

	Social Networks Only Affect	
	(1)	(2)
	Migration Cost	Job Arrival Rate
Social Network	-0.0045***	0.0021**
	(0.0005)	(0.0008)
Years of Education	0.0204***	0.0195***
	(0.0001)	(0.0002)
Age	0.1310***	0.1250***
	(0.000115)	(0.000191)
Age^2	-0.0019***	-0.0018***
	(0.0000)	(0.0000)
Type 1	0	0
	(.)	(.)
Type 2	-0.0563***	-0.0810***
	(0.000462)	(0.000733)
Type 3	-1.131***	-1.148***
	(0.000649)	(0.00165)
Constant	5.256***	5.311***
	(0.00226)	(0.00366)

1. Column 1 provides the correlate between social network status and rural migrants' log earnings when social network only affect migration costs not job arrival rate.
2. Column 2 provides the correlate between social network status and rural migrants' log earnings when social network only affect job arrival rate not migration costs.
3. Standard errors are presented in parentheses
4. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

between social network status and migrants' earnings. Column 2 shows the case when social network only affects job arrival rates, and in this situation, we expect a positive correlation. From the point estimates in Table 5, we find that social networks affect migrants' earnings as the model described, which also illustrates why the empirical analysis may have different findings of how social networks affect migrants' earnings.

Table 6 gives the decomposition results of the social network investment behaviours. The benchmark column provides the simulation results based on the estimates from the model. Individuals' investment decisions are affected by their locations and employment states. For example, when social networks affect both migration costs and the job arrival rate (i.e., in the benchmark column), 62.9% of rural individuals invest in their social networks, and only 22.4% of employed rural migrants invest in social networks. The reason is that social networks only affect the continuation values when individuals are employed in urban areas. Therefore, employed rural migrants have less incentive to invest in their social networks.

Under two restricted specifications, we can find that individuals effectively use and invest in social networks to optimize their migration decisions and labor market outcomes. If social networks only affect migration costs, individuals who live in rural areas invest most in their social networks (i.e., 52.1%), and less than 5% of migrants choose to invest in their networks. The reason is that social networks affect migration costs directly, so rural residents have the most incentive to invest in their networks. Social networks only affect the continuation values after individuals migrate to urban areas. When social networks effect is only through the job arrival rate, the investment rates for migrants are quite similar, but the non-migrants' investment rate is lower than the benchmark (e.g., 50.2% vs. 62.9%).

Table 6 also shows how individuals' social network investment choices respond to the effect of social networks. When social networks only affect migration costs, about 40% of individuals will invest. When networks affect the job arrival rate, 43% of individuals will invest. More individuals choose to invest in their social networks since the impact is larger than the migration cost channel. When both channels work, 50% of individuals invest in their social networks.

5.3 Policy Simulations

Since the Chinese government has set to increase the urbanization rate to 60% by 2020,[28] I propose three different hypothetical policies, all of which will achieve this aim. The policies include providing monthly unemployment benefits for rural migrants in urban areas and two types of lump-sum subsidies for migration costs.

First, I need to calculate the increase in the fraction of rural individuals migrating to urban areas to achieve the government's goal. Based on the annual report of the National Bureau of Statistics of China (NBSC), an additional 118 million rural individuals will need to migrate to urban areas.[29] Since I only consider rural male individuals in my data, this translates to an additional 76 million rural males mi-

[28] This target had been achieved in 2020, but most provinces have changed the definition of rural and urban hukou status, therefore, the target is not exactly achieved by migration behaviors.

[29] This number is calculated based on total population in 2011 (i.e., $1347.7 \times 0.6 - 690.8 = 117.8$).

Table 6: The Responses of Social Network Investment

	Benchmark	Social Networks Only Affect	
		Migration Costs	Job Arrival Rate
In Rural Areas			
Individuals who invest	62.87%	52.10%	50.20%
Individuals who do not invest	37.13%	47.90%	49.80%
In Urban Areas			
Employed			
Individuals who invest	22.44%	4.96%	22.52%
Individuals who do not invest	77.56%	95.04%	77.48%
Unemployed			
Individuals who invest	50.25%	1.01%	51.27%
Individuals who do not invest	49.75%	98.99%	48.73%
Total			
Individuals who invest	50.60%	40.40%	43.35%
Individuals who do not invest	49.40%	59.60%	56.65%

1. The benchmark uses the estimates from the model which allows individuals to invest in their social networks.
2. The migration cost column gives the counterfactual results if social networks only reduce migration costs.
3. The job arrival rate column gives the counterfactual results if social networks only increase the job arrival rate.

grating.[30] Therefore, the fraction of total migrants will be about 50%.[31] When I simulate different policies, I target the 50% of rural men that migrate to urban areas in my data when I simulate the model.

Table 7 provides the simulation results for the policy counterfactual simulations in three specifications. The first one is that the government provides monthly unemployment benefits in urban areas for rural migrants. This policy will increase the value of living in urban areas and decrease the return migration. The second policy is an unconditional lump-sum subsidy for migration costs when rural individuals migrate to urban areas. The third policy provides a conditional lump-sum subsidy for migration costs if migrants have social networks in urban areas.[32]

In Table 7, the second and third columns give the specifications of three policies when allowing for individuals' social network investment. If the government implements any of these policies, the urbanization goal can be achieved. The value of 284 means that the government pays an average of 284 yuan of unemployment benefits in urban areas monthly; The value of 802 in the row of migration cost means the policy of a lump-sum subsidy for average migration costs (802 yuan) per person if rural individuals migrate to urban areas. The value of 782 in the conditional migration costs row denotes the policy of providing a lump-sum subsidy for migrants who have social networks (782 yuan) per person. Under these three policies, the goal of the urbanization rate can be achieved. I can then compare these policies in terms of government budgets. The government will spend less to implement both policies of migration cost subsidy than the unemployment benefit policy to reach the goal of urbanization rate (155.73 (unconditional subsidy), 141.86 (conditional subsidy) vs. 396.53 (unemployment benefit) billion yuan). Since the conditional subsidy policy

[30] This number is calculated by the total additional migrants times the fraction of male migrants (i.e., $118 \times 64.5\% = 76$). 64.5% is the fraction of male migrants. Since NBSC does not provide the number of rural migrants, I use the number of rural men who did not migrate in 2011 divided by the fraction of rural individuals who reside in rural areas to calculate the number of rural men with rural registration (i.e., $\frac{250.59}{0.71} = 352.95$). I then calculate the number of male migrants (age 15-64) by subtracting the number of rural men who do not migrate (i.e., $352.95 - 250.59 = 102.36$ million). China Yearbook Rural Household Survey states that the total number of rural migrants in 2011 was 158.6 million. Therefore, the fraction of migrants who were male is $102.36/158.6 = 64.5\%$.

[31] The fraction is the sum of current rural male migrants and the additional rural males who will migrate divided by the total number of individuals with rural household registration (i.e., (76.0+102.4)/353.0=50.5%).

[32] There exists a similar migration policy in Canada. If the individual has relatives or family members in Canada, he will more easily pass the immigration requirements from the Canadian government.

Table 7: Policy Simulation Results: Government Budget

	With Network Investment		Without Network Investment	
	Policy (Yuan)	Government Budget (Billion Yuan)	Policy (Yuan)	Government Budget (Billion Yuan)
Unemployment Benefit	284	396.53	329	483.65
Unconditional Migration Cost	802	155.73	809	167.35
Conditional Migration Cost	782	141.86	775	147.04

1. In the model with network investment, individuals invest in their networks by sending gifts to their friends or relatives.
2. In the model without network investment, social networks are formed by a dynamic probit process.
3. This table provides the three different policies' simulation results, all of which achieve the goal of urbanization rate (i.e., 60% by 2020)
4. The unemployment benefits row shows the monthly value of unemployment benefits the government provides and the total budget the government pays to achieve the goal.
5. The unconditional migration cost row gives the value of a lump sum subsidy for migration costs if rural individuals migrate and the total budget the government pays.
6. The conditional migration cost row gives the value of a lump sum subsidy for migration costs if rural individuals with social networks in urban areas migrate and the total budget the government pays.

encourages individuals to invest in their social networks, the government will spend less than the unconditional migration cost subsidy.[33]

Table 7 also provides the policy simulations for the model without social network investment decisions. Under this case, I set the coefficient of network investment equals to zero (i.e., $\beta_1^s = 0$). This means that investment in social networks does not increase future network status. Therefore, no one will choose to invest in their networks. Comparing the models with and without social network investment decisions, we find that the government needs to spend more to attract rural people to migrate in the model with network investment decisions. The reason is that without network investment, individuals cannot increase the probability of having a network. Therefore, if individuals want to migrate, they need to pay more migration costs and have a lower job arrival rate. To offset the impact of network investment, the government has to spend more (i.e., 483.65 vs. 396.53 for unemployment benefits; 167.35 vs. 155.73 for unconditional migration cost subsides; 147.04 vs. 141.86 for conditional migration cost subsidies).

[33] Yuan is Chinese currency, which is equal to about 1/8 U.S. Dollar in 2000.

Table 8 shows the moments of earnings and choices before and after introducing the government policies. The effects of policies on rural migrants' earnings are different. For example, the unemployment benefit policy increases the value of the unemployment state and therefore increases their reservation earnings. As a result, the average earnings for rural migrants are higher compared to the average urban earnings under the other two policies. Also, the fraction of unemployed increases to 24% in urban areas under this policy, and the job search duration increases to 2.6 months. As a result, almost one-fourth of rural migrants are unemployed, causing the government to have to pay much more to achieve the urbanization goal.

The column of unconditional migration costs gives the individuals' choices under the policy of subsidy for migration costs. First, the average migration costs are significantly reduced by the government lump-sum subsidy. After introducing this policy, the actual average migration costs paid by migrants are around 100 yuan (i.e., 912-802=100). The individuals who are constrained by high migration costs are more likely to be affected by this policy. Second, providing a lump-sum subsidy for migration costs does not generate a large return and repeated migration behaviours. After this policy, the average migration times are 0.55 and 0.53 under two policies that subsidize migration costs and 0.38 for unemployment benefit policy. One of the reasons to explain this phenomenon is that there is no subsidy for return migration costs. Another reason is that the value of staying in urban areas is large. The fraction of employed in urban areas shows 90% of rural migrants have jobs, indicating why rural migrants do not have much return and repeated circular migrations.

The fifth column gives the simulation results if the government provides the conditional lump-sum subsidy for migration costs. Under this policy, only the migrants with social networks in urban areas can get the subsidy. This policy encourages rural individuals to invest in their social networks. Average migration costs paid by migrants are around 138 Yuan (i.e., 920-782=138). Since more rural migrants have social networks, the job search duration is shorter, and the fraction of employed migrants is similar to the policy of providing an unconditional migration cost subsidy. The government pays least under the policy of offering a conditional migration cost subsidy.[34]

The last raw compares the annual social network investment rates between different government policies with the original economy. We can find that the government

[34] From Table 7, the government budget for conditional migration cost subsidy is 142 billion, and the other two policies budgets are 156 and 397 billion.

policies are lower the individuals' incentive to invest in their social networks. It explains why the government would spend more to offset the impact of lower social network investment to achieve the same government target. When comparing the unconditional lump-sum subsidy policy for migration costs to conditional subsidy policy, more individuals invest in their networks to respond to the conditional lump-sum subsidy policy (e.g., 45% vs. 35%). It is also consistent with the finding in Table 7. It shows that the government budget is 141.86 billion yuan under the conditional lump-sum subsidy policy, which is less than the budget for unconditional lump-sum subsidy. Since the government policies crowd out individual investment in social networks, a policy that keeps the investment incentive can achieve the same government target at a lower cost.

6 Conclusion

This paper studies how social networks affect individuals' migration decisions and subsequent labor market outcomes while considering individuals' unobserved heterogeneity. I develop and estimate a dynamic model of circular migration that allows for a return and repeated migration.

To distinguish the effects of social networks through two different channels, I allow for the presence of a social network to directly affect migration costs and an indirect effect on labor outcomes via an impact on the job arrival rate. In the model, individuals can invest in their social networks to increase the probability of creating or sustaining them.

I use the Chinese Household Income Projects (2007-2009) panel data and estimate the model by the maximum likelihood method. The estimation results show that social networks affect individuals' migration choices and subsequent labor market outcomes through two channels: reducing migration costs and search frictions. Social networks reduce about 7% of migration costs and increase the job arrival rate for rural migrants by 41%. Unobserved types play different roles in individuals' outcomes. Type 2 individuals possess the talents to earn more, and type 3 individuals have lower migration costs and higher job arrival rates.

Table 8: Policy Simulation Results: Choices

	Before Policy	After Policy		
	Benchmark	Unemployment Benefits	Migration Costs (Unconditional)	Migration Cost (Conditional)
Urban Log Earnings	7.13	7.36	7.32	7.32
Rural Log Earnings	5.17	5.20	5.23	5.20
Fraction of Migrants	0.29	0.50	0.50	0.50
Fraction of Employees in Urban Areas	92%	76%	90%	90%
Fraction of Unemployed in Urban Areas	8%	24%	10%	10%
Job Search Duration (Month)	1.83	2.58	1.80	1.79
Average Job Arrival Rate	0.41	0.42	0.41	0.42
Average Migration Costs (Without Subsidy)	719	945	912	920
Average Migration Times (Times/Person)		0.38	0.55	0.53
Annual Social Network Investment Rate	50.60%	45.96%	35.22%	44.93%

1. The benchmark column shows the values simulated by the model without proposing the government policies.

2. This table provides the three different policies' simulation results, all of which achieve the same goal of urbanization rate (i.e., 60% by 2020)

3. Columns 3-5 give the simulation results when introducing the monthly unemployment benefits, an unconditional lump-sum subsidy for migration costs, and a conditional lump-sum subsidy for migration costs.

4. Earnings are the mean log earnings.

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

6. Average migration costs stand for the migration costs paid by the individuals before getting the government subsidies. The unit of migration costs is 1 Yuan.

7. Average migration times equals the total number of migrations after implementing the government policies divided by the total number of data sample.

The decomposition exercises show that social networks affect individuals' migration behavior more through the reducing search frictions channel than through the channel which lowers migration costs. For example, if social networks only reduce migration costs, 17% of individuals will migrate; if social networks only increase the job arrival rate, 27% will migrate.

The decomposition results also display that individuals effectively use and invest in social networks to optimize their migration decisions and labor market outcomes. Most individuals who invest in their social networks live in rural areas. Those who invest the most in their social networks in urban areas are the unemployed.

Next, I propose three different policies, all to meet a 60% urbanization rate by 2020. The policy simulations show that a migration cost subsidy policy will cost less than a policy providing unemployment benefits in urban areas. When comparing the two models (with and without social network investment decisions), I find that if individuals are allowed to invest in their social networks, they can effectively respond to the status of their social networks and invest in them to increase or keep them. To offset the individuals' responses, the government has to spend more to encourage rural people to migrate to urban areas. These results show that it is important to consider the different roles of social networks when studying migration decisions and policies intended to affect migration levels.

References

Barry R. Chiswick, Yew Liang Lee and Paul W. Miller. 2005. "Family Matters: the Role of the Family in Immigrants' Destination Language Acquisition." *Journal of Population Economics* 18 (4):631–647.

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

Brandt, Loren and Carsten Holz. 2006. "Spatial Price Differences in China: Estimates and Implications." *Economic Development and Cultural Change* 55 (1):43–86.

Buchinsky, Moshe, Chemi Gotlibovski, and Osnat Lifshitz. 2014. "Residential Location, Work Location, and Labor Market Outcomes of Immigrants in Israel." *Economatrica* 82 (3):995–1054.

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

Chan, Kam Wing. 2012. "Migration and Development in China: Trends, Geography and Current Issues." *Migration and Development* 1 (2):187–205.

Chandrasekhar, Arun, Melanie Morten, and Alessandra Peter. 2020. "Network-Based Hiring: Local Benefits; Gloabal Costs." *Working Paper* .

Colussi, Aldo. 2006. "Migrants' Networks:An Estimable Model fo Illegal Mexican Immigration." Working Paper.

Edin, Per-Anders, Peter Fredriksson, and Olof Åslund. 2003. "Ethnic Enclaves and the Economic Success Of Immigrants-Evidence From a Natural Experiment." *The Quarterly Journal of Economics* 118 (1):329–357.

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

Giles, John and Kyeongwon Yoo. 2007. "Precautionary Behavior, Migrant Networks, and Household Consumption Decisions: An Empirical Analysis Using Household Panel Data from Rural China." *The Review of Economics and Statistics* 89 (3):534–551.

Goel, Deepti and Kevin Lang. 2019. "Social Ties and the Job Search of Recent Immigrants." *ILR Review* 72 (2):355–381. Working Paper.

Hare, Denise and Shukai Zhao. 2000. *Lobor Migration as a Rural Development Strategy: a View from the Migration Origin*, chap. 6. Institute of East Asian Studies Research Papers and Policy Studies Series. University of California, Berkeley.

Harris, John R. and Michael P. Todaro. 1970. "Migration, Unemployment and Development: A Two-Sector Analysis." *American Economic Review* 60 (1):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 (2):271–320.

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* 76 (4):648–672.

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

Kennan and Walker. 2011. "The Effect of Expected Income on Individual Migration Decisions." *Econometrica* 79 (1):211–251.

Kono, Hisaki. 2006. "Employment with Connections: Negative Network Effects." *Journal of Development Economics* 81 (1):244–258.

Lewis, W. Arthur. 1954. "Economic Development with Unlimited Supplies of Labour." *The Manchester School* 22 (2):139–191.

Liang, Zai. 2007. *Chapter 12 Internal Migration: Policy Changes, Recent Trends and New Challenges*. New York: Oxford Unveristy Press.

Massey, Douglas S. and Kristin E. Espinosa. 1997. "What's Driving Mexico-U.S. Migration? A Theoretical, Empirical, and Policy Analysis." *American Journal of Sociology* 102 (4):939–999.

Meghir, Costas, Ahmed Mushfiq Mobarak, Corina Mommaerts, and Melanie Morten. 2020. "Migration and Informal Insurance." *forthcoming Review of Economic Studies* .

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

Montgomery, James Douglas. 1991. "Social Networks and Labor-Market Outcomes: Toward an Economic Analysis." *American Economic Review* 81 (5):1408–1418.

Morten, Melanie. 2019. "Temporary Migration and Endogenous Risk-Sharing in Village India." *Journal of Political Economy* 127 (1):1–46.

Munshi, Kaivan. 2003. "Networks in the Modern Economy: Mexican Migrants in the U.S. Labor Market." *Quarterly Journal of Economics* 118 (2):549–599.

———. 2020. "Social Networks and Migration." *Annual Review of Economics* .

Munshi, Kaivan and Mark Rosenzweig. 2016. "Networks and Misallocation: Insurance, Migration, and the Rural-Urban Wage Gap." *American Economic Review* 106 (1):46–98.

Navarro, Salvador and Jin Zhou. 2017. "Identifying Agent's Information Sets: an Application to a Lifecycle Model of Schooling, Consumption and Labor Supply." *Review of Economic Dynamics* .

———. 2020. "Human Capital and Migrarion: a Cautionary Tale." *Revised and Resubmitted at Journal of Econometrics* .

Sjaastad, Larry A. 1962. "The Costs and Returns of Human Migration." *Journal of Political Economy* 70 (5):80–93.

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

Zhang, Junfu and Zhong Zhao. 2015. "Social-Family Network and Self- Employment: Evidence from Temporary Rural-Urban Migrants in China." *IZA Journal of Labor and Development* 4 (4):1–21.

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