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

Jin Zhou

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

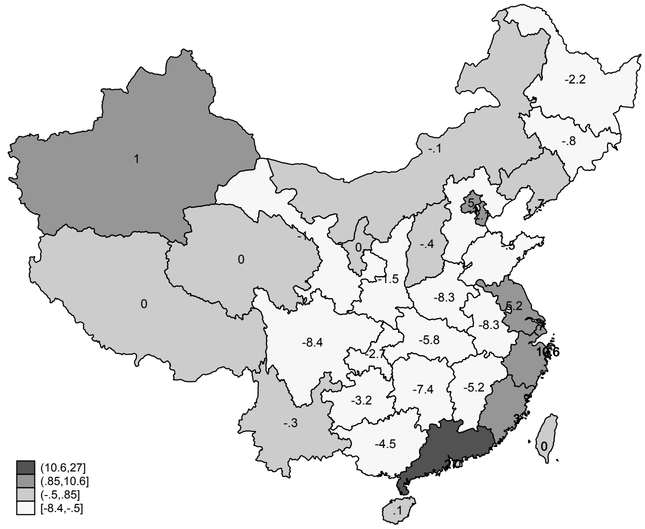
*I am grateful to Salvador Navarro, Nirav Mehta, Audra Bowlus, and Terry Sicular for their guidance. I thank James Heckman, Steven Durlauf, Lance Lochner, Joseph Altonji, Jaap Abbring, Hanming Fang, Paul Sullivan, John Ham, Chris Robinson, Todd Stinebrickner, and David Rivers for their helpful comments. 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.

A Background of Internal Migration in China

Table A1: Interprovincial Migration in China, 1990-2005 (In Thousands)

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

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



The numbers are the net percentage points of total migration in China from 2000 to 2005.

Figure A1: Migration Flows in China from 2000 to 2005

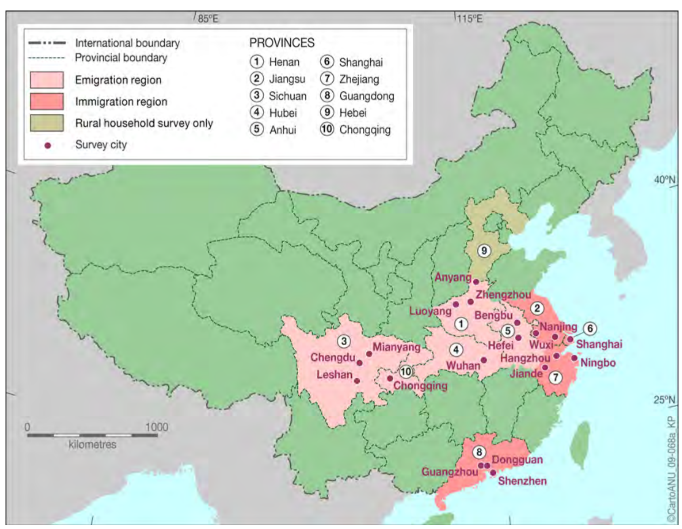


Figure A2: CHIP Data Sample Provinces in China

Table A2: Survival Analysis of the Age of First Time Migration

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

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

B Reduced Form Studies on the Role of Social Networks on Labor Market Outcomes

Table B1: The Relationship between Social Networks and Migration Decisions

Dependent Variable: Living in Urban Areas		
	OLS	Probit
Networks	0.0678	0.0420
	(0.0059)	(0.0060)
Education Year	-0.0045	-0.0026
	(0.0013)	(0.0013)
Married	0.0261	0.0012
	(0.0102)	(0.0095)
Number of Children	-0.0141	-0.0117
	(0.0034)	(0.0037)
Age	-0.0209	0.0019
	(0.0020)	(0.0021)
$\text{Age}^2 \times 100$	0.0037	-0.0242
	(0.0024)	(0.0026)
Constant	1.0771	0.3323
	(0.0370)	(0.0025)

1. The variable of social networks is the presence of social networks.
2. Married is the indicator variable for marital status.
3. The column of the probit model reports average marginal effects.
4. Standard errors are presented in parentheses.

Table B2: The Effects of Social Networks on Employment in Urban Areas

Dependent Variable: Employment		
	OLS	Probit
Networks	0.0279	0.0262
	(0.0021)	(0.0020)
Education Year	0.0013	0.0014
	(0.0005)	(0.0005)
Married	0.0116	0.0137
	(0.0028)	(0.0029)
Number of Children	-0.0095	-0.0099
	(0.0015)	(0.0092)
Age	0.0010	0.0009
	(0.0001)	(0.0001)
$\text{Age}^2 \times 100$	-0.0001	-0.0001
	(0.0000)	(0.0000)
Constant	0.6916	0.9151
	(0.0129)	(0.0009)

1. The variable of social networks is the presence of social networks.
2. Married is the indicator variable for marital status.
3. Standard errors are presented in parentheses.

Table B3: The Relationship between Social Networks and Network Investment

Dependent Variable: The Presence of Social Networks				
	OLS	2SLS	Probit	Probit with IV
Invest_{t-1}	0.0018	0.0151	0.0020	0.0358
	(0.0005)	(0.0040)	(0.0005)	(0.0134)
Education Year	0.0005	0.0005	0.0005	0.0007
	(0.0001)	(0.0001)	(0.0001)	(0.0002)
Networks_{t-1}	0.9732	0.9725	0.1126	0.1841
	(0.0005)	(0.0006)	(0.0019)	(0.0303)
Married	0.0033	0.0029	0.0037	0.0049
	(0.0007)	(0.0007)	(0.0007)	(0.0013)
Number of Children	-0.0012	-0.0011	-0.0013	-0.0019
	(0.0003)	(0.0003)	(0.0003)	(0.0006)
Constant	0.0144	0.0061	0.7606	0.7580
	(0.0012)	(0.0028)	(0.0002)	(0.0016)

1. The variable of social network investment is whether the individual send gifts to their friends or relatives at period $t-1$.
2. Married is the indicator variable for marital status.
3. The instruments for the 2SLS and probit models are the distance between rural counties where the individual lives to Beijing, Shanghai, and Guangzhou, which are the top three cities in China.
4. The coefficients for the probit models (with and without instruments) are average marginal effects.
5. Standard errors are presented in parentheses.

C Simple Model and Testing

C.1 The Simple Model

The framework is a basic discrete time search model. The individual has an infinite time horizon and makes decisions based on his locations. I will describe the decision process as follows:

If the individual is in a rural area, his wage ω is drawn from the distribution $G(\omega)$. He knows his migration cost M and the value of unemployment V^n. Hence, the migration decision is made based on the following equation:

$$Mig = \begin{cases} 1 & \text{if} \quad V^r \leq V^n - M \\ 0 & \text{else} \end{cases} \tag{1}$$

where $V^r = \frac{\omega}{\rho}$ is the value of staying at rural places.

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

1. If he just arrived in urban places, he has to be in the unemployment state for one period and pays a search cost ξ.
2. If the agent has already stayed for one period in the state of unemployment in urban places, he can get a job offer with probability λ, which follows a Poisson process.
   (a) If he gets a job offer(with probability λ), he makes a decision between three choices: unemployment in urban places, employment, or return migration (i.e., $\max\{V^n, V^e, V^r\}$).
   (b) If he does not get a job offer, (with probability $1 - \lambda$), he will select between two choices: unemployment in urban or return migration (i.e., $\max\{V^n, V^r\}$).
3. For simplicity, I do not consider on-the-job search. Hence the individual holds his job until the exogenous separation happens. Once the separation happens, with probability δ, he chooses between unemployment in urban cities or return migration (i.e., $\max\{V^n, V^r\}$).

The Bellman equations for each of the three states are[1]:

$$\begin{aligned}
V^r &= \frac{\omega}{\rho} \\
V^n &= -\xi + \frac{\lambda}{1+\rho} E_{\omega,w} \max\{V^{e'}, V^{n'}, V^{r'}\} + \frac{1-\lambda}{1+\rho} E_{\omega} \max\{V^{n'}, V^{r'}\} \\
V^e &= w + \frac{\delta}{1+\rho} E_{\omega} \max\{V^{n'}, V^{r'}\} + \frac{1-\delta}{1+\rho} E(V^{e'})
\end{aligned}$$

The value of staying at rural places V^r reflects that once the agent draws a wage ω from wage distribution $G(\omega)$, he takes that wage forever. Once the rural agent does not migrate at the initial period, he will not migrate in future. Thus the value of staying at rural places is $\frac{\omega}{\rho}$.[2]

If the unemployed migrant receives a wage offer, he compares three values: the value of the offer, the value of continuing to search, and the value of return migration. If the migrant does not receive a job offer, he compares the value of continued unemployment with the value of return migration. From the Bellman equation, we find that if the expected value of unemployment exceeds the present value of return migration, migrants will continue to search for a job. Once the wage offer is high enough (i.e., the value of employment is larger than the value of unemployment and the value of return migration), migrants will accept the job offer.

To document the mechanism of how social networks affect migration decisions and accepted wages, I include two extensions to the basic model. These two cases describe the effects of social networks on job arrival rates and migration costs, respectively.

Definition (Reservation Values) At reservation wage R, the individual is indifferent between employment at wage R and the better state between unemployment and return migration. i.e. $V^e(R) = \max\{V^n, V^r\}$. The value $V^e(R)$ is denoted as the reservation value.

Proposition 1

[1]Since I assume the model is stationary $\lambda, \delta, \xi, M, G(\omega)$, and $F(w)$ do not dependent on unemployment duration.

[2]This setting is a simplified version, and a more formal version is documented in the main content of the paper. The simplified model is used to illustrate the mechanisms of social networks on labor market outcomes.

(a) If the individuals with social networks have higher job arrival rates than the ones without (i.e., $\lambda(S=1) > \lambda(S=0)$), migrants with social networks have a higher mean reservation value than those without social networks. i.e., $E_{V^r,V^{r'}}(V^e(R)|Mig=1,S=0) < E_{V^r,V^{r'}}(V^e(R)|Mig=1,S=1)$

(b) If the individuals with social networks have lower migration costs (i.e., $M(S=1) < M(S=0)$), migrants with social networks have a lower mean reservation value than those without social networks. i.e. $E_{V^r,V^{r'}}(V^e(R)|Mig=1,S=1) < E_{V^r,V^{r'}}(V^e(R)|Mig=1,S=0)$

C.2 Case One: Social Networks only Affect Job Arrival Rates

In this model, social networks only affect the job arrival rate λ, (i.e., $\lambda(S=1) > \lambda(S=0)$). In particular, rural migrants with social networks have higher job arrival rates than those without social networks. Since the value function of unemployment is given by:

$$V^n = -\xi + \frac{\lambda(S)}{1+\rho} E_{w,\omega} \max\{V^{e'}, V^{n'}, V^{r'}\} + \frac{1-\lambda(S)}{1+\rho} E_{\omega} \max\{V^{n'}, V^{r'}\} \qquad (2)$$

It can be shown that V^n is strictly increasing in λ, which means $V^n(S=1) > V^n(S=0)$. Since $V^e(R) = \max\{V^n, V^r\}$, we can easily get the results that $E_\omega(V^e(R)|Mig=1,S=1) > E_\omega(V^e(R)|Mig=1,S=0)$, where ω is rural wage which is drawn from wage distribution $G(\cdot)$ and S is the social network indicator.

When social networks affect job arrival rates, the distributions of unemployment values between migrants with and without social networks differ. However, the cut-off value of unemployment in urban cities is the same between migrants with social networks and those without social networks. That is because social networks do not affect migration costs here. Since the distributions of unemployment are different between the group with social networks and the group without social networks, in this case, the model predicts that social networks affect wages even when not controlling for the migration selection.

If we let the reservation wage be denoted as $w_{it}^R = x'\beta + \gamma S_{it} + u_{it}$ for each individual and accepted wage w st. $w \geq w^R$. In random search framework, I assume that the offered wage w satisfies $w = w^R + \varepsilon_{it}$, where ε_{it} can be treated as a luck, so ε is independent of w^R. Only the accepted wage can be observed, which means that we only observe the wage conditional on $\varepsilon_{it} \geq 0$. Therefore, the migration decision is made as follows:

$$Mig = \begin{cases} 1 & \text{if} \quad V^r \leq V^n(S) - M \\ 0 & \text{else} \end{cases}$$

the migrants' reservation wage satisfies the condition:

$$w^R = x'\beta + \gamma S_{it} + u_{it} \geq a \tag{3}$$

where a includes the rural wage and migration costs, both of which are assumed to be constant.

The mean of migrants' accepted wages can be described by the following equation:

$$\begin{aligned} E(w|Mig = 1, x, S) &= x'\beta + \gamma S_{it} + E(u_{it}|M = 1, x, S) + E(\varepsilon|Mig = 1, x, S) \quad (4) \\ &= x'\beta + \gamma S_{it} + E(u_{it} 1_{u_{it} \geq a - x'\beta - S_{it}}) + E(\varepsilon|Mig = 1, x, S) \end{aligned}$$

From Equation (4), it is clear that even if controlling for the selection induced by migration (i.e., $E(u_{it} 1_{u_{it} \geq a - x'\beta - S_{it}})$), social networks still affect accepted wages. When I assume social networks only affect job arrival rates, the effect of social networks on wages should be positive even after controlling for the self-selection due to the migration decision.

C.3 Case Two: Social Networks only Affect Migration Cost

In this case, social networks only affect migration costs. (i.e., $(M(S = 1) < M(S = 0))$), and the migration decision is made as follows

$$Mig = \begin{cases} 1 & \text{if} \quad V^r \leq V^n - M(S) \\ 0 & \text{else} \end{cases}$$

The migration cost can be observed by the agent but cannot be observed by econometricians. I also assume that the agent with social networks has a smaller migration cost than the one without social networks, (i.e., $(M(S=1) < M(S=0))$).

Proposition 1 shows that migrants with social networks have a lower mean reservation wage than those without social networks. Since the distribution of unemployment has not been affected by social networks, migrants with social networks and those without social networks have the same distribution of unemployment. Thus, it can be proved that migrants with social networks have a lower mean of accepted wages than those without social networks due to the lower migration costs for the migrants with networks.

Since, in this case, social networks only affect migration costs, distributions of the value of unemployment are the same between migrants with social networks and without social networks. However, differential migration costs induced by social networks can generate different cut-off values of unemployment. The cut-off values are the lowest values for which rural individuals would migrate. As we discuss above, social networks only affect migration costs. Thus, networks do not affect reservation wages.

The migrant's reservation wage satisfies the following condition:

$$w_{it}^R = x'\beta + u_{it} \geq b + M(S) \tag{5}$$

where b is the strictly increasing function of rural wages. Similarly, I assume the offered wage satisfies $w_{it} = w_{it}^R + \varepsilon_{it}$, where ε_{it} can be treated as a luck. The migrant's accepted wage is given by $w_{it} \geq w_{it}^R$, where ε is independent of w^R. Hence, the mean of accepted wages is given by:

$$\begin{aligned} E(w|Mig=1,x,S) &= x'\beta + E(u_{it}|M=1,x,S) + E(\varepsilon|Mig=1,x,S) \\ &= x'\beta + E(u_{it}1_{u_{it}\geq b-x'\beta+M(S_{it})}) + E(\varepsilon|Mig=1,x,S) \end{aligned} \tag{6}$$

Equation (6) shows that once controlling for the selection induced by migration, social networks should not affect wages ($\gamma = 0$). This is because social networks affect wages only through the influence of migration costs.

C.4 Testing Model prediction

In order to test the differential predictions of these two mechanisms based on the data. I conduct the tests as follows:

$$\ln w_{it} = x_{it}'\beta + \gamma S_{it} + \pi\varphi_{it} + \nu_{it} \tag{7}$$

,where φ_{it} is the selection term of migration.

If social networks only affect job arrival rates (i.e., $\lambda(S=1) > \lambda(S=0)$, the model predicts that the coefficient of social networks in Equation (7), γ, should be significantly positive even after controlling for selection. Under case 1, the model predicts that agents with social networks have a higher mean accepted wage than those without social networks. In case 2, if social networks only affect migration costs (i.e., $M(S=1) < M(S=0)$), the coefficient of social networks in Equation (7),γ, should be insignificant after controlling for the selection term φ_{it}.

To summarize the models' predictions:

1 If I do not control for selection, and if the coefficient of social networks on wages is positive, the results are consistent with case 1. If the coefficient of social networks on wages is negative, the results are consistent with case 2.

2 After controlling for selection induced by migration, case 1 predicts the coefficient of social networks being positive, and case 2 predicts the coefficient of social networks being insignificant.

In Table C1, I conduct the tests using the Heckman two-step model. In the first step, I use a probit model of migration decision to calculate the selection term φ_{it} (the inverse Mill's ratio). In the second step, I estimate the Equation (7) with and without controlling for the selection term φ_{it}.

From Table C1, the first two columns show that the coefficients of social networks on wages are positive and statistically significant when the selection term is not included. This result is consistent with the prediction of case 1. However, after controlling for the selection term, the coefficients of social networks become insignificant. These results are consistent with the predictions of case 2 but not those of case 1.

The testing results show that we should consider both channels. It is hard to distinguish which channel dominates from the reduced form analysis. The formal structural model helps to answer the above question.

Table C1: The Tests of Mechanisms of Social Networks

Dependent log hourly earnings (With/Without Heckman Two-step Selection)				
	Without Correction		With Correction	
Network	0.0732***	0.0712***	0.0077	0.0098
	(0.0111)	(0.0109)	(0.0248)	(0.0243)
Years of education	0.0316***	0.0278***	0.0501***	0.0443***
	(0.0021)	(0.0021)	(0.0053)	(0.0053)
Age	0.0332***	0.0295***	0.0387***	0.0358***
	(0.0032)	(0.0032)	(0.0038)	(0.0037)
Age^2	−0.0004***	−0.0004***	−0.0002***	−0.0002***
	(0.0000)	(0.0000)	(0.0001)	(0.0001)
Male	0.1625***	0.1506***	0.0706***	0.0632***
	(0.0106)	(0.0107)	(0.0263)	(0.0259)
Work Experience	0.0107***	0.0108***	0.0128***	0.0127***
	(0.0010)	(0.0010)	(0.0009)	(0.0009)
Married	-0.0049	0.0010***	0.0410**	0.0506***
	(0.0153)	(0.0150)	(0.0175)	(0.0171)
Num of child	−0.0243***	−0.0240***	−0.0402***	−0.0382***
	(0.0071)	(0.0070)	(0.0089)	(0.0086)
Selection φ	No	No	−0.6243***	−0.5891***
			(0.1617)	(0.1595)
Industry	No	Yes	No	Yes
Occupation	No	Yes	No	Yes

1. $*$ significant at 10% level, $**$ at 5% level and $***$ at 1% level.

D Estimates

Table D1: The Estimates of Marriage and Fertility Transition Process

	Marriage		Fertility
Education (β_1^{ma})	0.0027	Education (β_5^f)	-0.0062
	(0.0007)		(0.0029)
Constant (β_0^{ma})	3.1671	Age (β_1^f)	0.2709
	(0.0062)		(0.0098)
lnγ	-2.4694	Age^2 (β_2^f)	-0.0050
	(0.0096)		(0.0002)
		Num of Children$_{t-1}$ (β_3^f)	-1.1181
			(0.0200)
		Num of Children$_{t-1}^2$ (β_4^f)	0.197
			(0.0068)
		Married (β_6^f)	1.9646
			(0.0248)
		Constant (β_0^f)	-6.0428
			(0.1435)

1. The variable of education is education year.
2. Married is the indicator variable for marital status.
3. The first two columns give the estimates of the survival analysis of marital transition.
4. Column 4 gives the estimates of fertility transition which is formed by a dynamic probit model.
5. Numbers in parentheses are standard errors.

Table D2: Panel A: Structural Model Estimates

Earnings Equation(Urban)		Social Network Probit Equation	
edu year	0.0232	marriage	0.0456
	(0.0000)		(0.0013)
expu	0.0036	num of children	-0.0334
	(0.0000)		(0.0004)
expr	0.0037	sn_{t-1}	3.6925
	(0.0000)		(0.0031)
$\text{expu}^2 \times 100$	-0.0021	inv_{t-1}	0.3075
	(0.0000)		(0.0004)
$\text{expr}^2 \times 100$	-0.0006	constant	-1.6802
	(0.0000)		(0.0019)
constant	6.8811	Psychic Value of Living in Rural Areas	
	(0.0001)	age	0.0112
Earnings Equation(Rural)			(0.0000)
		$\text{age}^2 \times 100$	0.0518
edu year	0.0270		(0.0000)
	(0.0000)	marriage	0.6385
expu	0.0012		(0.0001)
	(0.0000)	num of children	0.1001
expr	0.0044		(0.0000)
	(0.0000)	constant	-0.1386
$\text{expu}^2 \times 100$	-0.0064		(0.0001)
	(0.0000)	Job Arrival Rate	
$\text{expr}^2 \times 100$	-0.0005	social network	0.5621
	(0.0000)		(0.0004)
constant	4.1351	edu year	0.0999
	(0.0001)		(0.0001)
Unemployment Value		constant	-1.6224
			(0.0004)
marriage	0.2006	Job Destruction Rate	
	(0.0004)		
num of children	0.7620	edu year	0.0007
	(0.0001)		(0.0003)
age	0.0006	constant	-3.5648
	(0.0000)		(0.0017)
$\text{age}^2 \times 100$	-0.0048	Return Migration Cost	
	(0.0000)		
constant	0.4582	marriage	0.6662
	(0.0004)		(0.0015)
Migration Cost		num of children	0.1721
			(0.0005)
social network	-0.5466	cohort	-0.0156
	(0.0005)		(0.0002)
marriage	0.2494	constant	0.8259
	(0.0008)		(0.0014)
num of children	0.0416	Immediate Offer Probability	
	(0.0005)	social network	-0.2040
cohort	-0.0594		(0.0003)
	(0.0002)	edu year	0.0126
constant	7.8235		(0.0001)
	(0.0020)	constant	-0.2515
			(0.0007)

1. sn_{it} is an indicator variable that takes value 1 if the individual has social networks.
2. expu and expr stand for work experience in urban and rural areas, respectively. They are both measured in months. Age is measured in years.
3. Cohort is defined by the birth year-1999.
4. Marriage is an indicator of marital status that takes value 1 if the individual is married.
6. Numbers in parentheses are standard errors.

Table D3: Panel B: Structural Model Estimates

Intercept	Type 1	Type 2	Type 3
Log-wage			
Urban	6.881	6.885	6.208
	(0.000)	(0.000)	(0.000)
Rural	4.135	5.318	2.908
	(0.000)	(0.000)	(0.000)
Unemployment value	0.458	0.515	0.566
	(0.000)	(0.001)	(0.001)
Network equation	-1.680	-1.642	-1.652
	(0.002)	(0.002)	(0.005)
Migration cost	7.824	11.575	7.934
	(0.002)	(0.003)	(0.009)
Return migration cost	0.826	1.047	1.303
	(0.001)	(0.002)	(0.009)
Psychic value of living in rural area	-0.139	-1.321	0.529
	(0.000)	(0.000)	(0.000)
Job arrival rate	-1.622	-1.624	-1.494
	(0.000)	(0.001)	(0.001)
Immediate offer probability	-0.252	-0.967	-0.247
	(0.001)	(0.001)	(0.001)
Type probability			
Edu year ≤ 9	0.476	0.391	0.124
Edu year > 9	0.407	0.484	0.109

1. Numbers in parentheses are standard errors.

Table D4: Unobserved Type Distribution
(Conditional on Education Year)

Edu Year	Type 1	Type 2	Type 3
0	58.00%	20.36%	21.64%
1	58.53%	23.89%	17.58%
2	57.21%	25.46%	17.33%
3	55.54%	28.08%	16.38%
4	53.85%	31.08%	15.07%
5	52.18%	33.01%	14.80%
6	50.69%	35.21%	14.10%
7	48.74%	37.25%	14.01%
8	46.67%	39.96%	13.37%
9	45.06%	42.57%	12.37%
10	43.33%	45.11%	11.56%
11	41.21%	47.55%	11.24%
12	39.72%	49.75%	10.54%
13	39.76%	50.39%	9.85%
14	37.80%	53.61%	8.59%
15	34.89%	55.56%	9.56%
16	32.24%	59.08%	8.68%

Table D5: Model Fit: Earnings

	Data	Model
Migrants:		
Log(Earnings)	7.128	7.134
sd(log(Earnings))	0.458	0.665
with Networks	7.132	7.137
sd(log(Earnings))	0.459	0.672
without Networks	7.117	7.127
sd(log(Earnings))	0.456	0.648
Non-migrants:		
Log(Earnings)	5.202	5.173
sd(log(Earnings))	0.968	1.036
with Networks	5.221	5.187
sd(Earnings)	0.955	1.037
without Networks	5.155	5.138
sd(Earnings)	0.999	1.033

1. Migrants include people who were born in rural areas and currently work in urban areas.
2. sd(log(earnings)) stands for the standard deviation of log earnings.

Table D6: Model Fits: Choices

	Data	Model
Fraction of Individuals with Networks	0.7224	0.7164
Fraction of Rural Individuals Living in Urban Areas	0.2902	0.2926
Fraction with Networks	0.2179	0.2120
Fraction without Network	0.0723	0.0807
Fraction of Return Migrants	0.0075	0.0046
Fraction with Networks	0.0054	0.0030
Fraction without Networks	0.0022	0.0015
Fraction of Rural Individuals Migrating in a Given Month	0.0076	0.0056
Fraction with Networks	0.0054	0.0046
Fraction without Networks	0.0022	0.0010
Fraction of Individuals Getting Immediate Offer upon Migrating	0.0038	0.0032
Fraction with Networks	0.0027	0.0026
Fraction without Networks	0.0011	0.0006
Fraction of Employed Migrants in Urban Areas	0.2646	0.2704
Fraction of Employed Migrants with Network in Urban Areas	0.2003	0.1955
Fraction of Employed Migrants without Network in Urban Areas	0.0643	0.0749
Fraction of Unemployed Migrants	0.0246	0.0222
Fraction of Unemployed Migrants with Network	0.0176	0.0164
Fraction of Unemployed Migrants without Network	0.0070	0.0058
Average Job Search Duration (Months)	2.1972	1.8256

1. The data column provides the moments calculated based on the observations during 2007-2009.
2. The numbers in the table are averages divided by the 36 months (2007-2009) observed in the data.