Agent-Based Modeling for Rural Migration Decision and Action in China*

by

Lingxin Hao
Department of Sociology
Johns Hopkins University
3400 N. Charles Street
Baltimore, MD 21218
Tel: (410) 516-4022
Fax: (410) 516-7590
Email: hao@jhu.edu

and

Charlie Mitchell
Department of Sociology
Johns Hopkins University
3400 N. Charles Street
Baltimore, MD 21218
Email: cmitch53@jhu.edu

March 2013

* Prepared to be presented at the 2013 Annual Meetings of Population Association of America, New Orleans, Louisiana, April.

Agent-Based Modeling for Rural Migration Decision and Action in China

Abstract

This study develops mathematical models and computational techniques based on the principles of agent-based modeling (ABM) to study rural migration decision and action. We establish a conceptual model on rural-to-urban interprovincial migration in China by modifying migration theories to fit the China reality. The framework emphasizes the role of capital from foreign direct investment (FDI), networks of rural migrants from the same origin province, and urban employment prospect in determining rural peoples' migration preference structure. At the same time, urban FDI employers fill job vacancies by matching potential migrants with desirable characteristics such as age, education, and migrant networks. Migratory actions take place for those matched potential migrants. This framework is tested using ABM for 60 months from January 1996 to December 2000. The initial condition in December 1995 was estimated from the empirical data of the 2000 census as well as provincial statistics. Our ABM features both micro-macro feedbacks and micro-meso interactions. The goal of the ABM is to generate an artificial society where the trends and patterns of rural-urban migration driven by our theoretical framework are observed and compared with the trends and patterns from the census. If the results are similar, the ABM provides strong evidence to support our theoretical framework.

Agent-Based Modeling for Rural Migration Decision and Action in China

Introduction

The contemporary rural migration in China, the largest human movement in the world's history, started in the mid 1980s and has been rapidly growing since the mid 1990s. The trend is expected to continue into the next decade. Yet little is known about the causes of rural-to-urban migration for two reasons. Theoretically, migration itself changes the chief causal factors such as urban-rural wage differentials and migrant networks. Empirically rural migrants, despite their large size, are hard to sample because the rigid household registration (*hukou*) system records rural migrants in their home village even when they have spread the urban areas nationwide. Both problems present a great challenge to explain rural-to-urban migration in China.

This paper seeks to overcome the theoretical difficulty with agent-based modeling (ABM). We capitalize on ABM's ability of micro-macro feedbacks such that the migrant flow in one time changes the wage differential and migrant network in the next. We also exploit ABM's advantage in modeling interactions to enable the matching between potential migrants and urban employers. We bring the empirical basis of our ABM upfront and set the initial condition of our ABM based on the China 2000 census data and provincial provincial statistics. As a first step of a larger project, this paper focuses on modeling the origins and their corresponding destinations of interprovincial migration from 1995 to 2000.

Objectives

This study will develop mathematic models and computational techniques from the principles of ABM to study rural migration decision and action. We first establish a conceptual model on rural migration in China by modifying theories on causes of migration to fit the China reality. The framework emphasizes the role of capital from foreign direct investment (FDI), migrant networks, and urban employment prospect in determining rural migration decision and action in China. Second, this framework is tested using ABM based on empirical data from the 2000 Census and provincial statistics. The initial year is 1995, in which the known or estimated information includes the distribution of rural-origin labor force in rural vs. urban areas of each province, arable land, FDI capital stock in the real term, and origin-specific migrant networks in each province's urban area. Annual updated information includes arable land and FDI capital stock as exogenous changes. Monthly internal updates are made for rural-origin labor force distribution as a consequence of individual migration actions simulated monthly from January 1996 to December 2000. These internal updated macro-states every month contribute to the changing wage differentials and migrant networks when specific pairs of origin and destination are considered. Thus these two primary causal factors of rural-to-urban migration are dynamic through cross-level feedbacks, fundamentally different from conventional statistical models without addressing the endogeneity problem as well as cutting-edge statistical models addressing the endogeneity problem with instrumental variables or unobserved heterogeneity. The ABM's goal is to generate an artificial society under the parsimony of a few theory-based rules. By comparing the origin-destination patterns and trends between the artificial society and the society we observed based on the 2000 census, we can suggest whether the theory-based rules are powerful or not.

In sum the central objective of this study is to apply migration theories to agent-based simulation of rural migration from 1996-2000. This study will pave the way for a larger project that will create a generative society from 1990 to 2010 based on three censuses' data and 20 years of aggregate statistics and then predict the patterns in the future 10 years.

The Conceptual Framework and Mathematical Models

A person in a rural area faces the decision making and action taking regarding moving and settling in a chosen urban destination, or staying in the origin. Migration theories suggest structural forces from rural vs. urban productions and migrant networks determine the attractiveness of a potential destination (Massey et al. 1993). Whether a potential migrant will be employed in an urban destination is determined by the desirability of this person in the eye of urban employers (Todaro and Maruszko 1987). Whether a person takes the action of migration will be determined by the matching of the attractiveness of a potential destination from the person's perspective and the desirability of this potential migrant in the employer's perspective.

The Attractiveness of Destinations. The difference in rural vs. urban production generates the wage differential between the origin and a destination. Drawing from the Cobb-Douglass production function and wage determination from economics, we can express urban-rural wage differentials as a function of urban and rural labor and capital input. Migration may generate benefits but also costs, which are multifaceted, including physical, cultural, and psychological. The geographic distance between the origin and a destination may capture the multifaceted costs (Jasso and Rosenzweig 1990). The cumulative causation thesis of migration states that a large share of emigrants from an origin creates a pressure on outmigration, and a large number of migrants in a destination from the same origin may help reduce the costs (Massey et al. 1993; Hao 2012). The attractiveness of a destination vis a vis the origin can be expressed in a mathematic form. Let P be the unit price of the product, Q the quantity of product output, W the nominal wage, L the labor input, R the rent for capital, K the capital input, A the total factor productivity, and α and β are the output elasticity for the labor input and the capital input, respectively. Under perfectly competitive (output and input) markets, the profit-maximization problem faced by (price-taking) farms/firms is

$$\max_{(L)} PQ - WL - RK \quad s.t. \quad Q = Q(L,K) = AL^{\alpha}K^{\beta}, \quad given\ P, W, R, and\ K$$

The first-order condition implies equality between the real wage and the marginal product of labor,

$$w \left(= \frac{W}{P}\right) = \frac{\partial Q}{\partial L} = \alpha A L^{\alpha-1} K^{\beta}.$$

Assuming that P, A, α, β vary between the agricultural and industrial sectors, the real wage differential (urban wage in area j' vs. rural wage in area j) can then be written as

$$\frac{w_{u,j'}}{w_{r,j}} = \frac{\alpha_u A_u L_{j'}^{\alpha_u - 1} K_{j'}^{\beta_u}}{\alpha_r A_r L_j^{\alpha_r - 1} K_j^{\beta_r}}. \qquad (1)$$

Hence the wage differential can be calculated based on empirical data on labor input and capital input and the estimated parameters of output elasticity α, β and the urban-rural total factor productivity ratio A_u / A_r.

The choice set of a potential rural migrant includes 31 options: outmigration to a specific destination j' ($j' = 1, \ldots, 30$ for the urban areas in each of the 30 provinces) or to stay in the origin rural area j. A potential migrant's demographic characteristics will alter the attractiveness. For example, younger (x_i) and better-educated (e_i) people are valued more in labor-intensive, export-oriented industries.

The attractiveness can be expressed as the probability $\pi_{i,j,j'}$ for individual i migrating from rural area j to urban area j' as a function of the real wage differential $w_{u,j'} / w_{r,j}$ and migration costs (captured by a distance function $d_{j,j'}^{\rho}$, a social influence function of emigration $\mapsto \eta_j \equiv 1 - L_j / L_{j0}$, and a cost-reduction function of migrant networks in a destination $\mapsto L_{j \to j'}$, adjusted for individual heterogeneity in age x_i and education e_i, is given by:

$$\pi_{i,j,j'} = \frac{m_j^u L_{j\to j'}^{\lambda} x_i^{\gamma} e_i^{\xi} d_{j,j'}^{\theta} (w_{u,j'} / w_{r,j})}{1 + \sum_{l'=1}^{30} m_j^u L_{j\to l'}^{\lambda} x_i^{\gamma} e_i^{\xi} d_{j,l'}^{\theta} (w_{u,l'} / w_{r,j})}, \quad j' = 1,2,...,30, \quad j' = 1,...,30 \qquad (2)$$

$$= \frac{m_j^u L_{j\to j'}^{\lambda} x_i^{\gamma} e_i^{\xi} d_{j,j'}^{\theta} (w_{u,j'} / w_{r,j})}{1 + m_j^u x_i^{\gamma} e_i^{\xi} \sum_{l'=1}^{30} (L_{j\to l'}^{\lambda} d_{j,l'}^{\theta} w_{u,l'} / w_{r,j})};$$

$$\pi_{i,j,j} = \frac{1}{1 + \sum_{l'=1}^{30} m_j^u L_{j\to l'}^{\lambda} x_i^{\gamma} e_i^{\xi} d_{j,l'}^{\theta} (w_{u,l'} / w_{r,j})} = \frac{1}{1 + m_j^u x_i^{\gamma} e_i^{\xi} \sum_{l'=1}^{30} L_{j\to l'}^{\lambda} d_{j,l'}^{\theta} (w_{u,l'} / w_{r,j})}.$$

Employment Prospects in Destinations. Rural migration in China is economic migration in nature. The continuously reducing arable land and the disproportional population growth (due to differential rural-urban fertility rate) impose double pressures on labor surplus in rural areas. The primary motivation of rural migration is to find employment in urban areas. Urban employers, primarily e labor-intensive, export-oriented manufacturing corporations with FDI capital, consider the human capital of potential employees (measured by age and education). In addition, employers attempt to reduced recruiting costs through network hiring (Waldinger 1996). Taking all migrant-hiring firm as a collective corporate action, the desirability of a potential migrant for this corporate actor can be expressed as a function of the potential migrant's age x_{ji}, education e_{ji}, and his networks in this destination $L_{j\to j'}$:

$$y_{j'ji} = f(x_{ji}, e_{ji}, L_{j\to j'}) \qquad (3)$$

Migration Decision and Action. A potential migrant will rank the attractiveness of all potential destinations. The destinations with a higher attractiveness score than the origin will be candidate destinations. Exactly which destination the potential migrant to settle will depend on whether the potential migrant has a high enough desirability score for the destination firms. Methods that can match a top candidate destination for a potential migrant and the desirability of workers for a destination will provide a solution for the potential migrant to move to which destination.

The Computational Procedure

We detail the computational procedure in three steps. Each addresses how micro-macro feedbacks and micro-meso interactions are implemented.

Attractiveness Scores. We will compute the attractiveness scores of destinations vis a vis the origin for each potential migrant. A potential migrant is defined as a person who is currently residing in a rural area and has never migrated out of the origin. With all terms on the right-hand-side of Equation 2 observed or estimated, we can obtain 31 attractiveness scores for each potential migrant and rank them on a descent order. Because only the provinces ranked above the migrant's home provinces are possible options, we truncate this list just above the home province and retain candidate destination provinces, which are further placed to groups of destination provinces with the attractiveness difference less than a small value ε, yielding List A with ranked maximum $m<31$ groups of candidate destination provinces.

Desirability Scores. We will compute each destination's desirability scores for the typology of age-education groups (6 five-year age groups for ages 15-44 and 4 levels of education) and the network size in this destination from another province. We rank them on a descent order of $6\times4\times31=744$ ranked desirability scores. Depending on the FDI capital stock and existing migrant labor force, the demand for new migrant workers can be calculated and allotted to the top half of the ranked desirability scores, assuming firms prefer above average workers. This yields List B for destination's desirability of workers.

Job-Worker Matching. The hiring decision of firms will follow the logic of the deferred acceptance algorithm (DAA) (Roth 2008), the result of which is considered as Pareto optimal and stable: i.e., no alternative matching will do better. Through eliminating less desirable migrants sequentially, the firms of a destination determine which peasants are recruited and these people will move to that destination.

The matching mechanism involves aligning the descendent ranking of the expected values of all peasants for potential destinations and the descendent ranking of the expected values of destinations for a sequence of all peasants, such that the firms in a destination get their most desirable workers (none of the non-matched peasants are more desirable) and, at the same time, no more attractive destinations would offer a peasant a job. The DAA is just so designed to yield this stable, optimal matching. The DAA maximizes the utilities of the firms in a destination, subject to the firms' labor demands, which equals the number of matches. At the same time, the DAA maximizes the utility of each peasant, subject to the peasant's expected value of migrating to a destination.

Specifically, the matching mechanism involves matching the attractiveness for a potential migrant and the desirability for the corporate actor of a destination province. A maximum of m iterations of matching mechanism will be activated. The first iteration begins with a randomly selected potential migrant (Wang as an example) and Wang's List A of ranked groups of candidate destination provinces. On the top of the first group is Beijing. Beijing assesses Wang's desirability score, which is to be matched with a score in Beijing's List B. Three outcomes are possible. (1) Wang moves to Beijing if a match is found.   (2) Wang moves to Shanghai, another destination province in the top group of candidate destination provinces if Wang's desirability score does not matches any score in Beijing's List B but it does match one in Shanghai's List B. (3) Wang is set aside and waits for the second iteration if none of the destination provinces in the List-A top group yield a match. The first iteration goes through every potential migrant with a non-empty List A. Effectively the first iteration has dealt with the List-A top group for all potential migrants.

The second iteration in the same month will deal with the second top group on List A for all remaining potential migrants (excluding those who moved during the first iteration). This will repeat until the n_i-th iteration is finished for the n_i-th group on any potential migrants' List A. At this point the agent-based simulation moves to the second month.

At the beginning of the second month (and all other subsequent months), the macrostates are updated, including the interpolated new value of FDI capital stock, the recalculated demand for migrant workers, the wage differentials, the social influence of emigration, and the size of migrant networks in destinations. All these macrostate changes require the recalculation of the attractiveness scores and desirability scores, and the remaking of List A and List B for another set of iterations performed in the second month. At the beginning of each year, the FDI capital stock is exogenously updated, the labor force size is update by the exit of the 45 year olds and the entrance of the 16 year olds.

The Setup of Initial Conditions

The ABM's initial conditions are strictly based on the empirical estimates from China's Census 2000. To construct the December 1995 labor force distribution across the rural and urban areas in each province, we use the census questions about individuals' hukou registration place if the residence place is not in hukou place, and the year of migration. Table 1 shows the comprehensive patterns of rural migration officially defined across county boundary as well as spatial expansion of cities that changes a peasant's residence type from rural to urban without a move.

As our theoretical framework address rural-to-urban migration driving by FDI-induced job opportunities, urban-rural wage differentials, and origin-province-based migration networks, it best explain interprovincial migration and has little explanatory power in explaining intra-provincial migration such as migration for marriage purposes. For this reason, this paper models the decision and action of interprovincial migration, i.e., the movement of peasants with rural hukou from the original rural area of a province to the urban area of another province. We confine the initial population to those who had never moved by the end of 1995 and those who had already move cross provincial boundary to the urban areas of another province. Given the large size of rural labor force, we also confine the young labor force as aged 16-44 in each year. The initial population includes those aged 11-44 in 1995, who became 16-49 years old in 2000. Each year the group older than 44 exits and the group aged 16 enters. Table 2 shows the initial population counts, and their age and education distributions by the rural and urban areas in each province.

Using provincial statistics on FDI flows in the nominal term and the investment price index, we estimated the FDI stock in the real term using the 1979 constant Chinese currency. These estimates can be seen in Table 3. We also input the empirical data on the origin province for rural migrants who had already moved to urban areas by 1995, which enables the calculation of migrant networks.

Results

Three types of results will be presented. First, the origin-destination patterns and trends will be compared between the generated society and the observed (empirically estimated) society from the Census data. Figures 1-3 are templates for such comparisons. Second, we show the micro dynamics of 4 individuals as they go through the 60 month ticks, with their attractiveness scores for all potential destinations, their position in the DAA of their top and second top choices of destinations, and whether they are matched to the urban destination jobs, and whether they eventually moved during the 60 months. Third, we show the DAA iterations of two destinations in the first month to illustrate its Plato optimal property.

Table 1. Frequency Distribution of Agricultural-Hukou Population in 2000: China 2000 Census

Migration Status	All Ages	Aged 16-64
Inter-Provincial	(3.4%)	(4.8%)
1. City destination	15,431	13,726
2. Town destination	6,337	5,816
3. Rural destination	8,144	7,155
Intra-Provincial	(5.1%)	(5.7%)
4. City destination	22,531	16,742
5. Town destination	11,516	7,458
6. Rural destination	10,934	7,759
7. Not yet moved, current rural residence	601,196	353,026
8. Moved with HK, any destination	109,914	91,712
9. Not yet moved, current urban residence	96,914	57,643
10. Undetermined	169	157
Total	883,086	561,194

Data source: China 2000 Census, the long-form sample representing 0.095% of the population.
Note: Labor force is defined as aged 16 - 64 and currently not enrolled in school. The numbers in the table can be taken as roughly in 1,000 persons.

Table 2. Initial Population Counts and Characteristics in December 1995

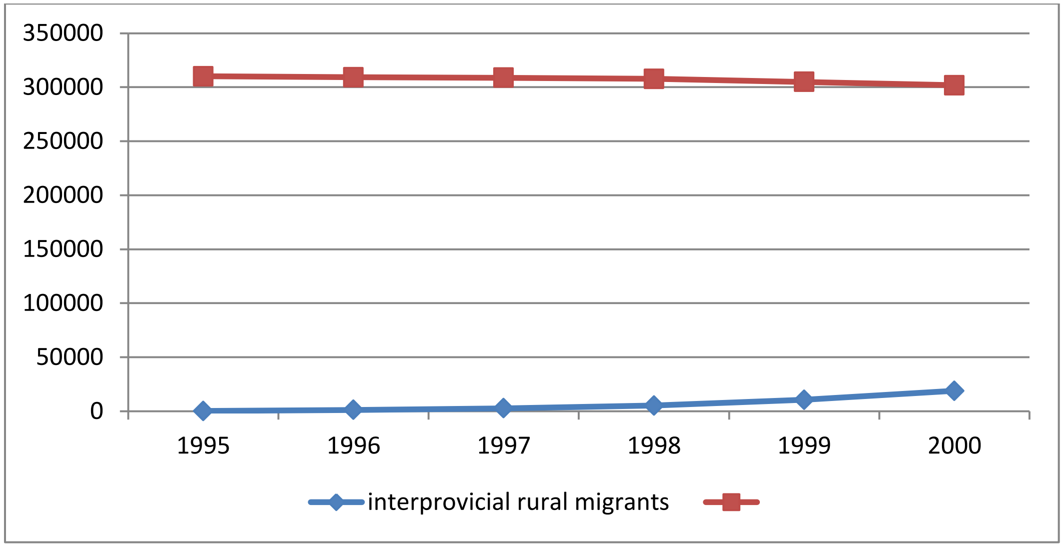
ru1995	seq	prov1995	pop	mcage1	mcage2	mcage3	mcage4	meds1	meds2	meds3
1	1	11	89	0.169	0.112	0.371	0.348	0.045	0.135	0.663
1	2	12	91	0.22	0.132	0.319	0.33	0.011	0.275	0.604
1	3	13	1990	0.252	0.143	0.33	0.275	0.024	0.245	0.651
1	4	14	794	0.243	0.174	0.32	0.263	0.028	0.27	0.636
1	5	15	492	0.272	0.159	0.313	0.256	0.12	0.339	0.48
1	6	21	739	0.225	0.137	0.336	0.303	0.041	0.341	0.568
1	7	22	533	0.238	0.154	0.341	0.266	0.071	0.396	0.484
1	8	23	648	0.269	0.176	0.323	0.233	0.068	0.355	0.531
1	9	31	62	0.145	0.129	0.323	0.403	0.081	0.21	0.581
1	10	32	1544	0.194	0.14	0.376	0.291	0.054	0.238	0.604
1	11	33	927	0.182	0.143	0.365	0.31	0.114	0.324	0.489
1	12	34	1766	0.245	0.181	0.374	0.2	0.104	0.326	0.526
1	13	35	745	0.252	0.166	0.326	0.255	0.129	0.345	0.472
1	14	36	1093	0.274	0.176	0.314	0.237	0.103	0.328	0.503
1	15	37	2332	0.216	0.143	0.364	0.278	0.057	0.287	0.586
1	16	41	2919	0.257	0.168	0.34	0.235	0.033	0.226	0.666
1	17	42	1328	0.224	0.15	0.367	0.26	0.088	0.317	0.516
1	18	43	1880	0.243	0.169	0.36	0.229	0.054	0.291	0.561
1	19	44	1036	0.232	0.135	0.345	0.289	0.067	0.317	0.542
1	20	45	1270	0.295	0.158	0.332	0.214	0.045	0.355	0.53
1	21	46	151	0.318	0.139	0.318	0.225	0.106	0.278	0.523
1	22	51	2956	0.197	0.203	0.348	0.252	0.094	0.419	0.442
1	23	52	1068	0.286	0.199	0.307	0.208	0.305	0.371	0.301
1	24	53	1344	0.282	0.186	0.331	0.201	0.243	0.463	0.269
1	25	54	82	0.366	0.171	0.293	0.171	0.817	0.171	0.012
1	26	61	896	0.21	0.175	0.34	0.275	0.098	0.289	0.528
1	27	62	721	0.222	0.187	0.384	0.207	0.237	0.343	0.358
1	28	63	120	0.325	0.2	0.292	0.183	0.442	0.275	0.25
1	29	64	124	0.363	0.169	0.274	0.194	0.266	0.274	0.403
1	30	65	410	0.346	0.193	0.278	0.183	0.117	0.4	0.41

2	1	11	13	0.231	0.154	0.538	0.077	0	0.231	0.538
2	2	12	4	0.25	0	0.25	0.5	0	0.25	0.75
2	3	13	1	0	0	1	0	0	0	1
2	4	14	1	0	1	0	0	0	0	1
2	5	15	0	0	0	0	0	0	0	0
2	6	21	5	0	0.4	0.6	0	0.2	0.4	0.2
2	7	22	1	0	0	0	1	0	0	1
2	8	23	2	0	0	1	0	0	0	1
2	9	31	24	0.167	0.5	0.208	0.125	0.042	0.167	0.625
2	10	32	9	0.111	0.333	0.444	0.111	0.111	0.333	0.556
2	11	33	13	0.308	0.231	0.385	0.077	0	0.231	0.769
2	12	34	2	0	0	0	1	0	0	1
2	13	35	5	0.2	0	0.6	0.2	0	0.4	0.4
2	14	36	0	0	0	0	0	0	0	0
2	15	37	3	0	0.333	0.667	0	0	0.333	0.667
2	16	41	2	0	0	0.5	0.5	0	0	1
2	17	42	0	0	0	0	0	0	0	0
2	18	43	1	0	0	1	0	0	1	0
2	19	44	105	0.286	0.305	0.352	0.057	0.029	0.086	0.705
2	20	45	4	0	0.25	0.75	0	0	0.25	0.5
2	21	46	2	0	0	0.5	0.5	0	0.5	0.5
2	22	51	0	0	0	0	0	0	0	0
2	23	52	2	0	0.5	0	0.5	0	0.5	0.5
2	24	53	9	0.111	0.222	0.667	0	0.111	0.333	0.444
2	25	54	3	0	0	0.333	0.667	0.333	0	0.667
2	26	61	1	0	0	1	0	0	0	1
2	27	62	0	0	0	0	0	0	0	0
2	28	63	0	0	0	0	0	0	0	0
2	29	64	2	0	0	1	0	0	0.5	0.5
2	30	65	9	0	0.444	0.556	0	0	0.222	0.778

Table 3. FDC Stock in the Real Term: 1995-2000.

provcode	fdi~1995	fdi~1996	fdi~1997	fdi~1998	fdi~1999	fdi~2000
11	287.895	354.532	414.384	495.144	578.023	664.809
12	104.114	154.681	214.832	270.146	321.181	368.245
13	49.187	76.71	107.672	140.967	165.4	176.167
14	7.297	12.334	22.012	29.992	43.311	48.292
15	10.136	10.995	12.789	14.574	16.251	18.281
21	66.811	84.434	107.205	127.531	143.834	164.844
22	55.414	73.262	88.288	92.452	92.574	94.033
23	37.316	48.062	61.75	71.061	78.625	85.456
31	349.935	471.544	583.27	647.834	688.589	729.235
32	346.519	448.628	560.953	688.221	799.38	898.616
33	81.307	107.17	130.248	147.676	168.939	189.9
34	22.095	29.948	35.387	37.311	40.767	43.006
35	169.985	209.781	246.583	281.154	310.651	334.118
36	17.216	21.667	29.124	35.407	38.361	39.116
37	231.534	289.921	339.365	375.943	416.969	467.576
41	79.195	104.648	129.732	149.068	162.764	175.434
42	44.829	57.175	68.64	85.827	89.509	93.357
43	26.252	37.497	51.473	61.779	67.97	73.804
44	687.902	869.014	1031.164	1189.05	1339.631	1475.974
45	49.662	59.376	72.689	84.854	91.033	93.923
46	59.648	67.211	72.775	78.328	79.578	79.744
51	90.578	93.78	100.75	115.245	121.993	127.135
52	6.432	7.876	9.759	12.272	14.838	17.079
53	12.308	15.258	17.508	19.076	20.674	21.486
54	0.436	0.55	0.511	0.57	0.559	0.505
61	30.517	36.685	49.554	52.479	54.203	56.334
62	16.207	19.66	20.619	20.729	20.924	22.13
63	1.151	1.396	2.33	3.662	3.602	5.479
64	4.063	4.686	5.133	5.678	5.875	6.907
65	4.929	6.044	6.035	5.946	5.928	5.783

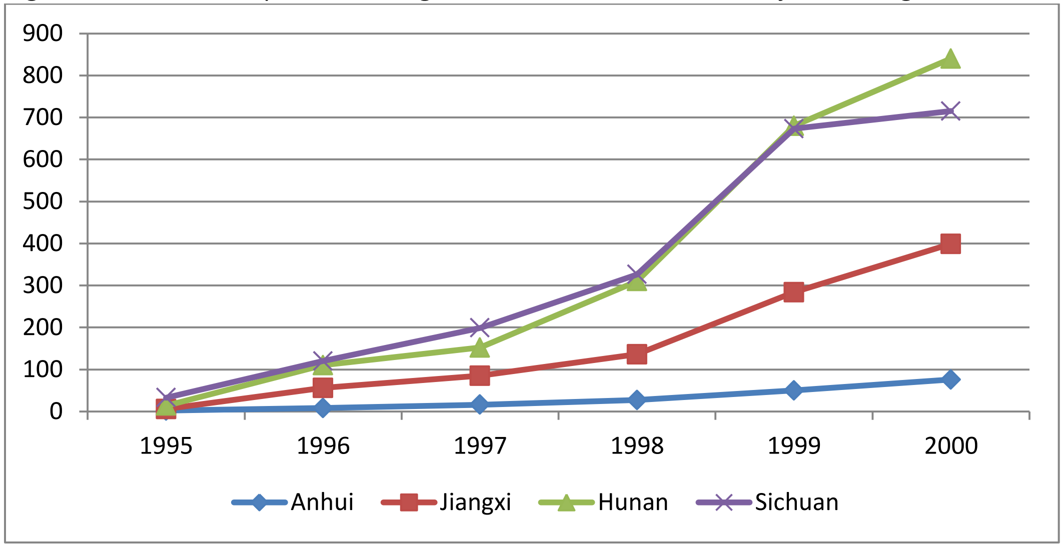
Figure 1. Cumulative Interprovincial Migration of Peasants in China: 1995-2000



(a) Census-observed

(b) ABM-generated

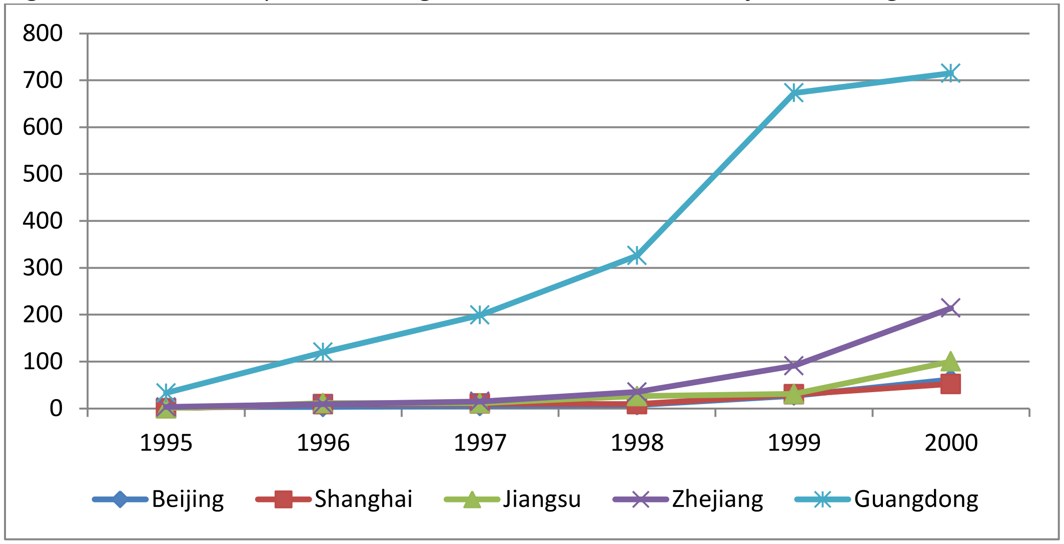
Figure 2. Annual Interprovincial Migration of Peasants from 4 Major Sending Provinces to Guangdong Province: 1996-2000



(a) Census-observed

(b) ABM-generated

Figure 3. Annual Interprovincial Migration of Peasants to 4 Major Receiving Provinces from Sichuan Province: 1996-2000



(a) Census-observed

(b) ABM-generated