Modeling China's Province-To-Province Migration Flows Using Spatial Interaction Model With Additional Variables

Jiaosheng He
Statistics Canada*

Jim Pooler
University of Saskatchewan**

The purpose of this paper is to model China's province-to-province migration flows using spatial interaction models. To date, conventional, gravity-type, spatial interaction models have typically employed information only on the number of movers and on the distances between places. This paper employs a new, multivariate approach to such interaction modeling by entering additional variables into the traditional modeling framework. In this particular application, two variables are added, namely, a measure of past migration, called migrant stock, and another measure of annual average total investment. The empirical verification of the models employs two interprovincial migration data sets for China. These involve the 1982–87 and 1985–90 migration flow data, consisting of two 28 x 28 data matrices. The results of the calibration show that the models with the additional variable(s) are capable of distributing migration flows with a much-improved degree of accuracy, in comparison with the conventional model. The calibration therefore provides empirical support for the validity of the multivariate approach to the spatial interaction modeling of migration flows.

Keywords: Spatial interaction model, additional variables, interprovincial migration flows, China.

The family of spatial interaction models (SIMs) due to Wilson (1967, 1971) has been regarded as a seminal work. It is indicated, however, that the SIMs can hardly provide behavioral explanation for migration processes because of the distributional nature of the model (Hay, 1991). In traditional gravity and Wilson-type models, the spatial separation, usually expressed as distance, time, or travel costs between regions, is the only variable for explanation (Fotheringham, 1997). Yet, the significance of distance or travel costs in spatial interaction, such as migration flows, has been reduced over the years (Plane, 1982; Chisholm, 1996). Although migration researchers (Snickars and Weibull, 1977; Webber, 1979) proposed a framework in which additional variables can be included in the conventional SIMs that may improve the model's

* Demography Division, Statistics Canada, Main Building, Tunney's Pasture, Ottawa, ON K1A 0T6 CA. E.mail: joe.he@statcan.ca
** Department of Geography, University of Saskatchewan, Saskatoon, S7N 5A5 CA. E.mail: jpooler@sasktel.net

explanatory power, the framework was rarely used, except for a few studies that employed an historical migration matrix (Plane, 1981; Pooler, 1985). The present research seeks to address this issue by incorporating two additional variables into conventional SIMs. Including variables at the destination is a process of weighing the destination's capabilities in attracting migrants from origin. The new models used in the present study are termed multivariate spatial interaction models (MSIMs). To this end, province-to-province Chinese migration data from two sources, the 1987 one percent survey and the 1990 national census, are employed.

Investigation of Chinese interprovincial migration flows has been very limited. Recently, there has been an attempt to estimate China's regional migration using the log-linear version of the classical gravity model (Shen, 1999). However, attempts to model province-to-province migration flows have been lacking. Hence, the present research is significant in two ways for Chinese migration studies. In the first place, migration modeling undertaken so far has been primarily based on concepts from Western free-market economies (Haynes and Fotheringham, 1984). Little attention has been given to migration in other types of nations in which a mixed economy with planned and free-market mechanism exists. Therefore, it may be useful and interesting to model migration flow patterns in a socioeconomic setting such as China's. In the second place, because the migration process is a response to socioeconomic differentials, it would be equally appealing to model interprovincial migration flows to provide some scientific guide for China's socioeconomic planning.

This paper is organized as follows. A brief introduction of the background of China's interprovincial migration is provided first. This is followed by the derivation of the MSIMs, a discussion of the two additional variables, the modeling results. and a conclusion.

BACKGROUND OF INTERPROVINCIAL MIGRATION IN CHINA

Prior to the late 1970s, internal migration in China was carefully restricted for the sake of social stability and for protecting the benefits of urban residents. In particular, rural-to-urban migration was strictly controlled (Goldstein, 1990). Migration restrictions had been exercised through the use of the Household Registration System (HRS) since the late 1950s and in particular since the end of the Great Leap Forward (1959–61). The HRS specifies that all who lived in rural areas, and were not state employees, were to be treated as agricultural householders and were not eligible for grain rations and other daily necessities from the State. Urban residents and state employees, however, were treated as members of urban or non-agricultural households and were eligible for guaranteed grain rations (Cheng and Selden, 1994). In 1958, the Chinese government stipulated the Ordinances of Household Registration for Chinese Residence, which required those who wished to move to urban places to have permission from the destination city (Zhang, 1988).

In the period before the late 1970s, some interprovincial migration had been allowed to take place. Such population movement, however, was not based on migrants' calculations of costs and returns, but rather depended upon socioeconomic planning strategy, or the military and ideological considerations of the government (Ma and Wei, 1997). One of the economic strategies was to develop the inland provinces in the 1960s and 1970s, called the *third-front construction*, which mobilized considerable population from coastal areas into the backward inland provinces (Yabuki, 1995; Linge and Forbes, 1990; Naughton, 1988). During approximately the same period of time, another program, the *urban-to-rural youth transfer*, also known as the youth rustication *(shang-shan xia-xiang),* was carried out. It was a resettlement scheme to transfer urban graduates of secondary schools to rural areas, in particular to the frontier regions such as Xinjiang (Northwest), Heilonjiang (Northeast), and Yunnan (Southwest) (Bernstein, 1977; White III, 1979; see Figure 1). It is estimated that the number of the dispatched urban youths is in the range between 13.2 million (White III, 1979) and 17 million (Shen and Tong, 1992) in the period of 1960s and 1970s. The geographical origins of the transfer were urban centers, in particular the three big cities of Beijing, Tianjin, and Shanghai.

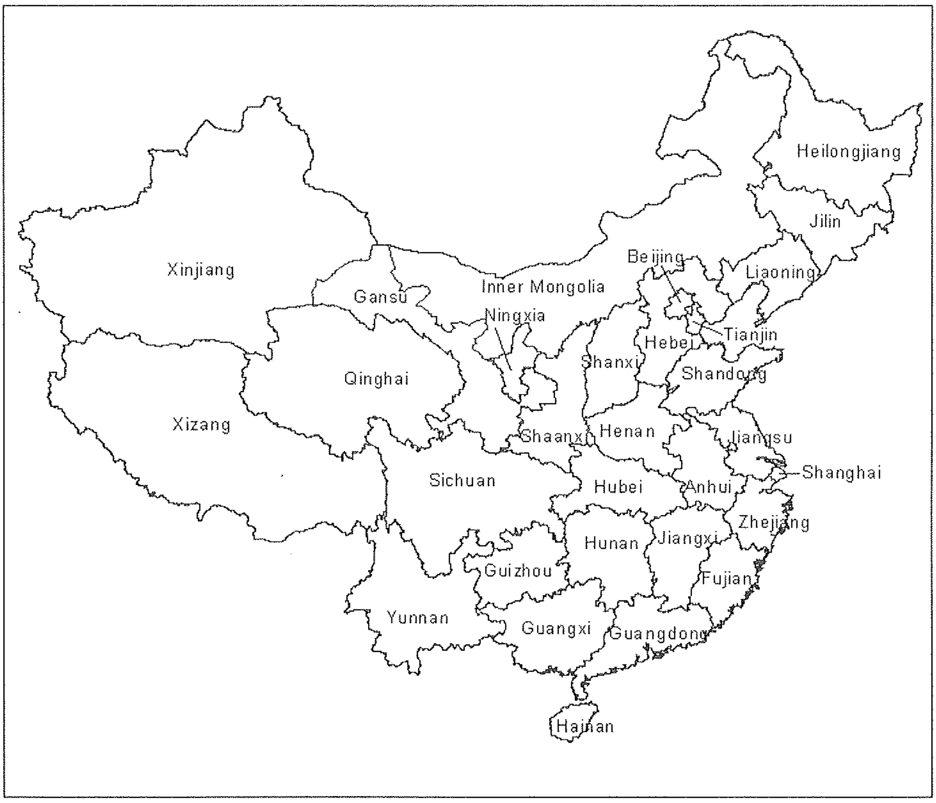
With the end of implementation of these two programs in the late 1970s, return migration to the origins in the coast provinces occurred in the 1980s due to the emphasis on the open door policy (Banister, 1987). The implication of such return flows is apparent since they constitute an important part of the interprovincial migration during the period 1982–1990.

Since the late 1970s the long-standing restricted migration policy in China has been relaxed with the introduction of economic reforms that aim at nurturing a free-market mechanism in the socialist economy. This policy change not only relieved some of the burden of rural population pressure, but also facilitated urban-to-urban migration. As a result, the volume of internal migration has greatly increased. However, information on the size of interprovincial migration flows was not available until the 1987 one percent population survey (China, 1988) and the 1990 national census (China, 1993).

The 1987 survey collected the information on inter- and intra-provincial migration in the period from 1982 to 1987. The survey tabulated interprovincial migrants as those who moved from cities, towns, or rural counties of other provinces to the place of enumeration and were living there on 1 July 1987. They consist of two types of migrants: (1) those who registered their official residence in the new place and; (2) those who did not, but had left their place of origin at least six months earlier, and had lived in the new place equal to or fewer than five years.

Migrants, based on the 1990 census definition, were referred to as those whose current place of residence on 1 July 1990 was in a different county or city on 1 July 1985. They included all those who changed their official residence registration to the new counties or cities, and also those who did not, but who had left their place of origin for at least one year and had lived in the new localities equal to or fewer than five years. Those who moved within a county or city were, therefore, not counted

Figure 1: Provinces of Mainland China, 1990.



as migrants (China, 1993(4):509, 512–513). Similar to the tabulations of the 1987 survey, the 1990 census distinguished between inter- and intra- provincial migrants. Interprovincial migrants were those who moved to the place of enumeration from cities, towns, and counties of other provinces. The total number of interprovincial migrants for China as a whole (28 provinces) increased from 6.24 million in 1982–87 to 10.75 million in 1985–90. During the 1982–87 period, Heilonjiang was the largest net loser (the number of net migrants was -0.26 million), while Shanghai had the largest net gain of 0.28 million migrants. Return migration flows were one of the major factors in migration patterns in this period.

During the 1985–90 period, however, Guangdong became the most important destination for interprovincial migrants, its net migrants being approximately one million. Sichuan was the largest net exporter of migrants with a net loss of 0.86 million. The overall pattern of interior-to-coast migration in 1985–90 was even clearer than in 1982–87.

Hainan, formerly part of Guangdong province, was given provincial status in 1988, and no data on migration to Tibet were collected. Therefore, theses two provinces are excluded from the analysis. Except for Tibet and Hainan (Figure 1), a 28 x 28 interprovincial migration data matrix can be derived from these two sources. The migration data from the 1990 census can further be disaggregated into male and female populations. Such a disaggregation will create a unique opportunity to make some empirical observations regarding male and female interprovincial migration in China.

THE MODEL

Suppose a known interprovincial migration matrix is available. This implies that outmigrants, O_i, leaving each origin, and inmigrants, D_j, arriving at each destination are known. The modeling problem is to estimate the probability, p_{ij}, that a migrant will move from origin to destination. The well-known entropy maximizing (EM) method is to maximize the Shannon entropy (Shannon and Weaver, 1949).

$$H = -\sum_{i,j} p_{ij} \ln p_{ij} \tag{1}$$

subject to constraints:

$$\sum_{j=1}^{n} p_{ij} M = O_i , \tag{2}$$

$$\sum_{i=1}^{m} p_{ij} M = D_j , \tag{3}$$

$$\sum_{i,j} p_{ij} d_{ij} = \bar{d} \tag{4}$$

where H is the Shannon entropy measure, M is total number of migrants in the whole migration matrix, d_{ij} is the distance between provinces, and $\bar{d}$ is the average or mean distance traveled by all migrants. Equations (2) and (3) are constraints that ensure that the *predicted* O_i and D_j are consistent with the *observed* O_i and D_j. Equation (4) is the constraint for the mean migration distance. The result of the maximization is

$$p_{ij} = \exp(-\lambda_i - \gamma_j - \beta d_{ij}) \tag{5}$$

which assigns migrants to provinces in the least biased way, subject to the constraints (2) to (4) (where λ_i, γ_j, and β are parameters associated with constraints (2), (3), and (4) respectively). When p_{ij} is defined with respect to interprovincial migration in the form of $p_{ij} = M_{ij}/M$, then Equation (5) can be expressed in a more familiar form of

$$M_{ij} = A_i\, O_i\, B_j\, D_j \exp(-\beta\, d_{ij}) \tag{6}$$

where M_{ij} is the predicted number of migrants between provinces, and $M_{ij} = p_{ij}M$. The balancing factors are represented by A_i and B_i (Fotheringham and O'Kelly, 1989; Pooler, 1994).

Suppose that in addition to the known information on migration above, further information on multiple prior probabilities related to migration are available, that is

$$\prod_k q_{ij}^{(k)} \quad (i, j = 1,\ldots n;\ k=1,\ldots m) \quad \text{kr} \tag{7}$$

These multiple prior probabilities are defined based on $q_{ij} = r_{ij} / \sum_{i,j} r_{ij}$, where r_{ij} is any measure on, or between provinces, which can be justified to have an effect on interprovincial migration. Minimizing the Kullback (1959) information

$$I(p:q) = \sum_{i,j} p_{ij} \ln \frac{p_{ij}}{\prod_k q_{ij}^{(k)}}, \tag{8}$$

subject to constraints (2), (3), (4), and (7) will produce a model that contains the multiple prior information q_{ij}. However, another adjustment is required in that constraint (4) is replaced by

$$\sum_{i,j} p_{ij} \ln d_{ij} = \overline{\ln d}, \tag{9}$$

Equation (9) is the constraint for the mean logarithmic distance. The rationale for using constraint (9) on the spatial separation is that it results in an inverse power function of distance deterrence for the model that can be more appropriate for modeling inter-regional or longer distance migration (Fotheringham and O'Kelly, 1989; Ottensmann, 1997). Such an inverse power function is employed in the present study. It is established that the mean trip length is used as the cost constraint if the cost function is exponential ($e^{-\beta d_{ij}}$), For the inverse power function, however, the term $d_{ij}^{-\beta}$ can be written as $e^{-\beta \ln d_{ij}}$, and thus the mean value of logarithmic distance (ln d_{ij} in Equation (9) is the cost constraint (Wilson, 1971, 1974). A number of researchers adopt the mean value of logarithmic distance (cost) to derive or to calibrate the spatial interaction models, such as Wilson (1974), Pooler (1994, 1995) and Fotheringham and O'Kelly (1989). When the inverse power function is used, the costs or spatial separation between regions can be defined as the logarithm of distance. Therefore, the interpretation of the inverse power function is that distance is perceived in logarithmic form by migrants.

The result of the minimization of (8), subject to constraints (2), (3), and (9), together with consideration of $M_{ij} = p_{ij}M$ is

$$M_{ij} = \prod_k q_{ij}^{(k)} A_i O_i B_j D_j d_{ij}^{-\beta} \tag{10}$$

Equation (10) is the multivariate spatial interaction model. It can be seen that Equation (10) is an estimation of migration flows, which allows for inclusion of k prior probability distributions. In the present study, a production- and cost-constrained version of the MSIM is used,

$$M_{ij} = \prod_k q_{ij}^{(k)} A_i O_i d_{ij}^{-\beta} \tag{11}$$

which results when Equation (8) is minimized, subject only to constraints (2) and (9). The balancing factor A_i ensures that constraint (2) is satisfied, and is defined as

$$A_i = \left[\sum_j \prod_k q_{ij}^{(k)} d_{ij}^{-\beta} \right]^{-1} \tag{12}$$

where $\prod_k q_{ij}^{(k)}$ in the present study includes the two additional explanatory variables, the migrant stock (s_{ij}) and the annual average total investment (h_j), which will be discussed in the next section.

When q_{ij} in Equation (11) is replaced by the two explanatory variables, Equation (11) becomes

$$M_{ij} = h_j s_{ij} A_i O_i d_{ij}^{-\beta} \tag{13}$$

where

$$A_i = \left[\sum_j h_j s_{ij} d_{ij}^{-\beta} \right]^{-1} \tag{14}$$

Equation (13) is the production- and cost-constrained MSIM used in the present study.

An important component in the process of modeling is to evaluate the model's ability to replicate known migration flows. A more accurate replication may demonstrate that the proposed model has a solid empirical basis, and can be used for prediction with confidence. On the other hand, less accuracy of the model may alert investigators to look into the causes behind it. Such causes are usually important clues for further improvement of the model.

Many goodness-of-fit statistics have been used in spatial interaction modeling. These statistics are discussed and reviewed by Wilson (1976), Knudsen and Fotheringham (1986), and Fotheringham and O'Kelly (1989). The present study employs the percent misallocated to assess the model's ability to replicate the known migration flows.

This statistic can be calculated using the predicted and observed migration flow matrices. It takes the following form:

$$\%E = \frac{50}{M} \sum_{i,j} \left| M_{ij} - M_{ij}^{*} \right|, \tag{15}$$

where $\%E$ represents the percentage of misallocated migrants in the migration matrix, M is total number of migrants, and $\left| M_{ij} - M_{ij}^{*} \right|$, is the absolute difference between the predicted and the observed migrant flows from origin to destination, respectively. Any calculated value for this measure indicates the percentage of migrants that would have to be redistributed to the correct provinces in order for the predicted migration matrix to match the observed matrix of migration. Based on this measure, further explorations can be made as to which provinces, and how many migrants, are over- or under-predicted.

TWO ADDITIONAL VARIABLES

Migrant Stock

Migrant stock refers to a group of people who previously moved to a destination, implying that an area of origin has a number of migrants who are already at the destination. Migrant stock is taken to represent the social networks that are of pivotal importance in labor migration (Massey et al., 1993; Massey, 1990, 1986; Montgomery, 1991). As mentioned earlier, China's migration prior to the late 1970s resulted mainly from political and military considerations. Two main interprovincial migrations occurred in the period of mid-1960s through the 1970s. The first was the sending of urban youths to the interior or frontier provinces, and the second was the migration caused by the relocation of industry from the coast to interior provinces (the so-called third-front industrial construction). A large number of urban youths were sent to the countryside and remote regions, and one to two million were engaged in interprovincial migration during the height of the Cultural Revolution in 1968–76 (Bernstein, 1977; Shen and Tong, 1992). Most of these transferred urban youth were allowed to return the areas of origin after 10 to 12 years (Banister, 1987). The interprovincial migration due to the construction of the 'three-fronts' industries between 1964 and 1979 (Shen and Tong, 1992) was also reversed in the 1980s. To take these effects into account in modeling Chinese migration, the variable of migrant stock is employed. Return migrants result from the changing political and economic situation in the 1980s. This implies that if a coast province has a large migrant stock in an interior or frontier province, this coast province is most likely to experience a large wave of return migration.

Also, it should be noted that not every migrant to the interior and frontier regions returned back to their home place, since some people married a local person, and in particular, some of the relocated industrial plants were in the interior permanently. In such cases, the migrant stock constitutes a major facilitator in attracting subsequent migrants from the places of origin. Therefore, migrant stock has particular significance in the modeling of interprovincial migration in China when compared with a migration system with a free market economy in the developed world. It should be noted that linking the migrant stock with the subsequent migration is not new (Plane, 1981; Pooler, 1995), though it has never attempted for Chinese migration data. What the present study emphasizes is that using the concept of the migrant stock can capture some unique features of the Chinese interprovincial migration flows due to political and ideological interventions.

Theoretical exploration on the roles played by chain migration has led to an eventual formulation of social network theory, which is increasingly used to explain migration (Massey et al., 1993). According to this theory, rural labor migrants may lack human or financial capital but possess some form of social capital (networks established in the community). It is upon this form of social capital that rural labor migrants heavily relied to facilitate their mobility strategies.

Given the fact that the formal communication mechanisms in developing countries, such as telephone and internet services, are usually not adequate, informal connections facilitated by migration-chain effects would have been widespread in rural-to-urban or inter-regional migration (Brown and Stetzer, 1996). In the 1980s, the migrant enclaves found in China's large cities, such as Beijing, Shanghai and Guangzhou, are clear indications of the overwhelming importance of the chain migration (Ma and Xiang, 1998). In addition, sample surveys conducted in the early 1990s revealed that 57 to 78 percent of rural labor migrants obtained employment through friends and relatives at the destination in China (Zhang et al., 1997). This demonstrates that the migrant stock who was voluntarily moved to prosperous areas in the reform period acts as an important facilitator to attract migrants from origins.

In the present study, migrant stock is defined as the number of people who moved previously to the destinations. In order for this measure to be fully operational, it can be expressed as a probability of the total number of migrants. Put in the form of the following equation.

$$s_{ij} = \frac{M_{ij}^{o}}{\sum_{i,j} M_{ij}^{o}}, \quad (i = 1, \ldots, n;\ j = 1, \ldots, n) \tag{16}$$

where s_{ij} is probability of migrant stock and M_{ij}^{o} is the number of people in the destination who moved from the origin during the period 1982–87.

Selection of Destination Variables

The section above has discussed the migrant stock variable in China's migration context, which is origin and destination specific. The question turns to what destination

Table 1: The correlation matrix for variables (1) to (6).

Variable	*(1)*	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*
(1)	1.000					
(2)	-0.269	1.000				
(3)	-0.223	0.936**	1.000			
(4)	0.518**	-0.187	-0.250	1.000		
(5)	-0.372	0.373	0.253	-0.041	1.000	
(6)	-0.019	0.841**	0.877**	0.093	0.184	1.000

Notes: Variable names are indicated in the text. ** correlation is significant at the 0.01 level (2-tailed).

variables should be selected and included in this particular migration modelling. One important rule for selection of variables in modelling migration is to avoid using inter-correlated variables (Alonso, 1968), or multicollinearity. Based on this rule, a correlation matrix is calculated among six potential destination variables, which are available to the present study (China, 1996). These six variables are (1) ratio of cultivated land to population in 1990, (2) per capita GDP in 1990, (3) per capita total investment in 1986–90, (4) length of railway in 1988, (5) annual average total investment in 1986–90, and (6) percentage of urban population in 1990.

If the correlation coefficients between variables are highly significant, one of each pair of the variables will be excluded from the models. On the basis of examination of the correlation matrix, as illustrated in Table 1, it is found that the correlation coefficients are highly significant among four of the pairs of variables, that is, variables (4)-(1), (3)-(2), (6)-(2), and (6)-(3). Thus, either variables (4), (3), (6) or variables (1), (2), (3) should be excluded.

A second rule for including destination variables into the migration models in the present study is the extent to which the variables have improved the performance of the model. In this study, the model is calibrated with variables (1) to (6), and the calibrated results are shown in Table 2. The model satisfies both origin and cost constraints. Preliminary calibrations of the model with variables (1) to (6) in Table 2 show that only variable (5), the annual average total investment, improves the model performance by 6.85 percent in comparison with the conventional model, and other variables offered no improvement in performance to the model. This may be a result of the fact that all the other variables cannot be justified as having significant effects on China's inter-provincial migration. As a result, according to the second rule, variable (5) remains in the model. On the other hand, as discussed in the preceding paragraph, variable (5) also has no significant correlation with other variables. Therefore, on the basis of the correlation analysis and preliminary calibration for the six variables, it may be concluded that variable (5), the annual average total investment in 1986–90, is properly included in the model.

Table 2: Model results for migration in 1985–90: percentage of migrants misallocated (% E).

Variable	*% E*	*Beta*
The conventional model	34.6339	1.1317
(1)	41.7926	1.2362
(2)	34.6340	1.1320
(3)	37.2584	1.1466
(4)	40.4040	1.2157
(5)	27.7793	1.1208
(6)	36.1392	1.1493

Note: Variable names are indicated in the text. The conventional model is $M_{ij} = A_i O_i d_{ij}^{-\beta}$. The calibrated model equation that includes possible destination variable is $M_{ij} = A_i O_i d_{ij}^{-\beta} w_j$, where w_j represents any one of the six variables as specified in the text, and all other terms are defined in section 3.

Annual Average Total Investment

In China, even though forces that determine investment patterns may be different from those in free market system (particularly so prior to the economic reform of the late 1970s), it is observed that the capital investment determines productive capacity, creates labor demand, and eventually prompts inter-regional migration. The investment discussed in this paper includes domestic loans, foreign investment, fund raising, investment from the central government's appropriation, and others. Two questions are worth discussion. The first is why the annual average of total investment is chosen rather than any specific yearly total investment, and the second is why total investment, rather than foreign investment, is chosen.

The amount of investment in China displays a cyclic pattern in the period prior to the economic reform (Ma and Wei, 1997). Yearly fluctuation of the total investment in the 1980s is also observed. This is because the investment depends on China's own economic cycle, and foreign investors are sensitive to policy changes and political stability. Therefore, the annual average of the investment can avoid such fluctuation. Specifically, the annual average total investment between 1986 and 1990 (five-year average) is used for modeling migration in the period 1985–90, while the annual average total investment between 1985 and 1987 (three-year average, data before 1985 is not available) is used for modeling migration in 1982–87. The data on the total annual investment are from the State Statistical Bureau of China (China, 1996).

Previous investigations on China's migration, together with studies of China's regional development, have used foreign investment data for explanation. One of the obvious shortcomings of using the foreign investment data is that other investments are ignored, and therefore a distorted picture may arise. Even for Guangdong, with the largest amount of the foreign investment among the provinces of China, the foreign investment is not the most important variable to explain migration at county level (Fan, 1996). Fan's observation may suggest that the foreign investment variable is not adequate enough to account for migration, since it excludes other types of investments. According to Kueh (1992), for each dollar of foreign investment received, an average of three *yuan* (Chinese currency unit) is spent for providing infrastructure, such as electricity, transport, and so on. Therefore, the foreign investment in China generates corresponding internal investments, and the internal investments create employment opportunities as well.

Another factor concerning China's investment is the power of fiscal transfer by the central government. Although China decentralized its fiscal system and allowed provinces to retain a considerable amount of revenue from local economic activities in the 1980s, fiscal transfer by the central government still has an impact on overall capital flows into some interior provinces and the border regions. Guangxi, Yunnan, Guizhou, Xinjiang, Qinghai, Ningxia, Tibet and Inner Mongolia all received subsidies. The fiscal transfer not only alleviates poor economic conditions, but also provides job chances. All these aspects of investment should be taken into account. The annual average total investment reflects the above factors of investments.

The variable of the annual average total investment, v_j, enters into the model in a probability form in order to avoid the effect of the units of the variable. It only changes the scaling of the balancing factors, but smaller balancing factors are more convenient to interpret. The variable is calculated as follows

$$h_j = \frac{v_j}{\sum_j v_j}, \quad (j = 1, \ldots n) \tag{17}$$

where h_j is probability of the annual average total investment in province j. The variable is calculated for two time periods, one for the period between 1985–87, and the other for 1986–90.

ANALYSIS OF THE MODELING RESULTS

In the actual modeling process, Equation (13) and its three disaggregated models, as specified in Equations (18) to (21), are estimated for comparative purposes.

$$M_{ij} = A_i O_i d_{ij}^{-\beta} \qquad \text{(CM)} \tag{18}$$

$$M_{ij} = h_j A_i O_i d_{ij}^{-\beta} \qquad \text{(MSIM1)} \tag{19}$$

$$M_{ij} = s_{ij} A_i O_i d_{ij}^{-\beta} \qquad \text{(MSIM2)} \qquad (20)$$

$$M_{ij} = h_j s_{ij} A_i O_i d_{ij}^{-\beta} \qquad \text{(MDIM3)} \qquad (21)$$

All the terms were introduced and defined earlier and are not repeated here. For convenience of presentation and analysis, Equations (18), (19), (20), and (21) are referred to as the conventional model (CM), MSIM1, MSIM2, and MSIM3, respectively, where MSIM refers to multivariate spatial interaction model.

Overall Performance, 1985–90 and 1982–87

Table 3 presents the modeling results for total interprovincial migration during the period 1985–1990. Three observations can be made from this table. First, the value for beta (β) decreases from 1.1317 for the CM to 0.4132 for MSIM3, indicating a 63.5 percent decrease. This change illustrates that when the two variables, the annual average total investment in 1986–90 and the migrant stock of 1982–87, are included, the effect of distance impedance on interprovincial migration appears to be partially replaced by these two variables. This reduction may imply that a destination province with high employment opportunities will attract a considerable number of migrants. For certain groups of migrants in some provinces, physical distance seems to be a secondary consideration for migration decision-making. For example, migration from Xinjiang to Guangdong is a case in point.

Second, a comparison of the beta values between MSIM1 and MSIM2 on the one hand, and the conventional model on the other, shows that the migrant stock variable in MSIM2 has a much more depressing effect on the beta value than the investment variable in MSIM1. This is an obvious confirmation that in a developing country like China past migration plays an important role in the process of interprovincial migration in terms of overcoming the difficulty of the distance barriers (Cai, 1999). It also might be that the migration stock variable partially replaces the distance variable since migration stock itself tends to be a negative function of physical distance. As distance of migration between origin and destination increases, the chance of having an effective social network becomes smaller. Therefore, in such circumstances distance and migration stock, to some degree, are complementary variables. Also, a certain

Table 3: Parameters of the CM and the MSIMs for total interprovincial migration, 1985–90.

Model type	*CM*	*MSIM1*	*MSIM2*	*MSIM3*
Beta	1.1317	1.1208	0.2263	0.4132
% *E*	34.63	29.78	16.73	17.76

number of migrants are driven by marriage. They move over long distances, or even across several provinces, for instance, those who migrate from southwestern China to the coastal provinces (Fan and Huang, 1998). Apparently, to a certain degree, the marriage migration in China helps explain the reduction in the value of beta.

Third, the goodness-of-fit statistic, represented by the percentage of migrants misallocated (shown as % *E* in Table 3), shows that the performance of the model is improved considerably from the conventional model (34. 63 percent) to MSIM1 (29.78 percent) to MSIM3 (17.76 percent).[1] The percentage of migrants misallocated among the cells of the migration matrix measures the percent of the total number of predicted migrants that are allocated to incorrect cells in the migration matrix. The magnitude of this index illustrates that the conventional model correctly assigns about 65 percent of the total number of migrants in 1985–90, whereas MSIM3, with inclusion of the two additional variables, correctly allocates about 82 percent of all migrants, this being a 17-percentage point improvement. This also illustrates the validity of the multivariate approach to modeling China's interprovincial migration.

Finally, Table 3 also indicates that the performance of all the models with additional variables(s) is better than the conventional model. Moreover, a comparison of all the models with additional variable(s) shows that MSIM2 performs best, albeit its performance is close to that of the MSIM3. This observation illustrates that for this particular study, the inclusion of additional variables into the spatial interaction model does not necessarily lead to improvement of performance. This somewhat counter-intuitive finding may be partially a result of a fitting criterion being only to let the estimated and observed mean of logarithmic migration distance converge.

For modeling the interprovincial migration in 1982–87, MSIMs 2 and 3 are not estimated, because the migrant stock variable is not available. Table 4 shows the modeling results. The table indicates that the beta value is in a trend of decrease, and the overall performance as measured by the percentage of migrants misallocated is improved by 4.69 percent between the CM and MSIM1. This again confirms the validity of the MSIM. It can be noted that the magnitude of such an improvement of the model performance is comparable with the improvement (4.85 percent) between the CM and MSIM1 for migration in 1985–90 as shown in Table 3.

Male and Female Migration

The seminal treatment of gender differentials in migration was by Ravenstein (1885). He noted that the principal characteristics of female migration can be summarized as

Table 4: Parameters of the CM and MSIM1, 1982–87.

Model type	*CM*	*MSIM1*
Beta	1.0851	1.0737
% *E*	38.60	33.91

follows: (1) women are more mobile than men, (2) women dominate short distance migration but men dominate long distance migration, and (3) like men, women migrate based on economic motivations. Ravenstein's observations provide an important basis on which research on the gender issues of migration in the contemporary era can be undertaken.

The relative numbers of men and women in migration streams can be measured by sex ratios (i.e., number of men per 100 women). Three patterns of gender differentials in migration have been identified in developing countries: high sex ratios in lower income countries of Africa, female dominance (or low sex ratios) in Latin America, and diverse sex ratios in Asia (Hugo, 1991).

In China, the sex ratios for out-migration ranged from a high of 286 in Beijing to a low of 47.5 in Guizhou in the 1985–90 period, with most provinces (24 or 25 out of 28 provinces) showing high sex ratios in both in- and out-migration flows (China, 1993). It can be assumed that female migrants in China were more hindered by the spatial distance and relied more on their own social networks (here previous migrants) than do the male migrants.

Table 5 summarizes statistics of the modeling results for both male and female interprovincial migrants in 1985–90. It appears from the table that a difference can be clearly identified with respect to properties of the estimated model between male and female migrations. The following paragraphs summarize the difference.

First, for the conventional model and MSIM1, the difference in the values of beta between male and female migrants indicates that female interprovincial migrants are more likely to be constrained by the spatial separation. However, for MSIMs 2 and 3 the values for beta for female interprovincial migration are considerably lower than those for the male migration, and this is particularly so for the MSIM2. Such a disparity in the values for beta indicates that in interprovincial migration in China females are more likely to follow their predecessors, illustrating the importance of the past migrants in guiding and channeling the subsequent migrations.

Second, with the addition of the two variables into the model, the beta values are in general in a trend of decrease from the CM to MSIM1 to MSIM3. For example, the beta value for male migration decreases from 1.1138 for the CM to 0.4081 for MSIM3, while the corresponding value for female migration declined from 1.1576 to 0.3287. A comparison of the beta values between MSIM1 and MSIM2 reveals that the two introduced variables exert a varied influence on male and female migrations.

Table 5: Parameters of the CM and MSIMs for male and female migrations, 1985–90.

Model type	*CM*		*MSIM1*		*MSIM2*		*MSIM3*	
	male	female	male	female	male	female	male	female
Beta	1.1138	1.1576	1.0836	1.1986	0.2243	0.1406	0.4081	0.3287
% *E*	32.83	39.50	29.59	32.08	17.55	20.62	18.39	19.79

Table 6: Rank correlation coefficient (r_S) between the balancing factor (A_i) for total migration,1985–90.

Models compared	r_S	*Significance (two-tailed)*
CM & MSIM1	0.950	0.000
CM & MSIM2	0.185	0.346
CM & MSIM3	0.373	0.050
MSIM1 & MSIM2	0.144	0.465
MSIM1 & MSIM3	0.337	0.080
MSIM2 & MSIM3	0.963	0.000

Finally, the performance of the estimated models, measured by the percentage of migrants misallocated, is better for male migration than that for female migration. However, a common feature is that the performance is improved from the CM to MSIM1 and to MSIM3. It should be noted that the main purpose of the inclusion of the MSIM2 in Table 5 is to illustrate whether the two variables have a different impact on replicating the observed migration flows, as compared to MSIM1. Table 5 indicates that such a varied impact does exist.

In order to evaluate the impact of the variables further, Table 6 shows the calculated rank correlation coefficients between the balancing factors of the migration models for total interprovincial migration in the period 1985–90. It is reasoned that when the impact of a variable is larger, the rank correlation between the two sets of the balancing factors arising from the two different models would tend to be smaller. It is indicated from the table that the rank correlation coefficient between the CM and MSIM1 is 0.950, while the corresponding value between the CM and MSIM2 is 0.185. This demonstrates that when the investment variable is entered into the model, the ranking of the balancing factors is similar to that for the CM, whereas when including the migrant stock variable in the CM, the rank of the balancing factor changes considerably. Again, this result leads to the conclusion that the impact of the migrant stock variable on the model is more considerable than the investment variable.

Error Analysis—Residuals

The most general way of identifying errors in the model predictions is to make a comparison between the observed and predicted migration flows. Although in the above discussion the percentage of misallocated migrants is used as the goodness-of-fit statistic to evaluate the overall accuracy in prediction of the models, the prediction errors for the specific province-to-province flows cannot be uncovered by this overall statistic. Therefore, a residual analysis between the observed and predicted migration flows is undertaken for this purpose.

The relative residuals are calculated for the purpose of investigating the accuracy of prediction of interprovincial migration between specific pairs of provinces. Following the convention indicated by Thomas (1968) and Thomas and Huggett (1980), the relative residuals can be expressed as follows.

$$RE = (M_{ij}^{*} - M_{ij}) / M_{ij}^{*}, \tag{22}$$

where *RE* represents the relative residual, and M_{ij}^{*} and M_{ij} are the observed and predicted migration flows, respectively. It is clear from Equation (22) that negative relative residuals imply over-predicted reality, while positive residuals represent under-prediction of the observed migration flows. This measure is useful in several ways. First, it can be easily calculated and interpreted. Second, the relative residuals calculated in this way for the present case could reveal particular errors in prediction for interprovincial migrations between specified pairs of provinces. A further aspect of the relative residuals is the possibility of the mapping of this measure or the examination of the residual values for migrants either leaving or entering a particular province based on the residual matrix. The pattern of residual signs provides information on the observed attractiveness of provinces for interprovincial migration flows. The power of such attractiveness can be found from the magnitude of the residual values.

It would be tedious and unnecessary to discuss the relative residuals associated with all the models calibrated in the present study. Rather, the relative residuals based on the CM and MSIM3 are discussed to demonstrate to what extent the predicted flows of the models match the observed flows. Table 7 shows the average of the relative residuals and standard deviations for out-migration, based on the CM and MSIM3. The average values are calculated based on all the specific residual values for a particular origin province. For example, the average value for Beijing is -0.485 based on the CM, and this figure is computed based on all of the 27 relative residuals for out-migration flows from Beijing to the 27 possible destination provinces. It is indicated from Table 7 that for the CM, for all the 28 origin provinces, the average of the relative residuals are negative, implying, on average or in most cases, that out-migration flows are over-predicted. The highest average value is found in the provinces of Zhejiang (-0.029), whereas the lowest one is in the provinces of Guangxi (-10.9). It appears in Table 7 that the average relative residuals are highly associated with their standard deviations. Indeed, the correlation coefficient between these two indices is found to be -0.987. Therefore, the average of the relative residuals based on the CM shows that the migration flows are poorly predicted by the model, or the spatial distance is not sufficient to explain the observed migration patterns. This corresponds with the results revealed by the goodness-of-fit statistic as discussed earlier.

Table 7: Average and standard deviation of the relative residuals for out-migration flows, 1985–90.

		CM		*MSIM3*	
ID no.	*Province*	*average*	*sd*	*average*	*sd*
1	Beijing	-0.485	1.801	-0.103	0.465
2	Tianjin	-0.670	1.297	-0.070	0.494
3	Hebei	-1.210	1.693	-0.460	0.698
4	Shanxi	-0.818	1.822	0.020	0.704
5	In. Mongolia	-3.233	5.683	-0.609	0.883
6	Liaoning	-0.847	1.273	-0.107	0.483
7	Jilin	-2.357	3.283	-0.201	0.441
8	Heilongjiang	-3.499	5.974	-0.430	0.662
9	Shanghai	-1.288	1.722	-0.320	0.784
10	Jiangsu	-0.642	1.071	-0.299	0.769
11	Zhejiang	-0.029	0.550	0.065	0.374
12	Anhui	-1.061	1.303	-0.246	0.649
13	Fujian	-1.331	2.436	-0.302	0.697
14	Jiangxi	-2.654	6.311	-0.411	0.794
15	Shandong	-0.683	1.092	-0.242	0.644
16	Henan	-0.660	1.152	-0.183	0.445
17	Hubei	-1.040	2.459	-0.238	1.073
18	Hunan	-1.825	2.813	-0.751	2.072
19	Guangdong	-2.226	5.471	-0.271	0.734
20	Guangxi	-10.900	26.267	-1.821	1.902
21	Sichuan	-1.077	2.482	-0.185	0.342
22	Guizhou	-6.673	15.920	-0.787	1.482
23	Yunnan	-6.899	16.972	-0.577	1.038
24	Shaanxi	-0.883	1.792	-0.043	0.379
25	Gansu	-1.878	5.356	-0.475	1.111
26	Qinghai	-1.487	2.790	-0.337	1.297
27	Ningxia	-2.450	6.459	-0.563	1.468
28	Xinjiang	-2.651	4.088	-0.348	0.554

For MSIM3, the average values and standard deviations of the relative residuals are reduced considerably in comparison with those derived from the CM. The average values of the relative residuals based on MSIM3 range from 0.02 in Shanxi to -1.821 in Guangxi. Only in one other province is the average of the relative residuals slightly larger than that of Shanxi; and this is Zhejiang, with the average value of 0.065. All other average values are negative. This observation demonstrates that the accuracy of the prediction of the interprovincial flows by MSIM3 is greatly improved. It also echoes, in general, the result revealed by the goodness-of-fit as already examined.

However, the average value of the relative residuals for out-migration flows cannot reveal the errors in prediction for interprovincial migration between *specific pairs* of provinces. It may also not indicate both the sign and magnitude of the residuals for specific flows. Hence, it is necessary to use other ways to investigate such aspects of the relative residuals. The residual matrix can be employed to fulfill such a purpose. Table 8 shows the matrix of the relative residuals based on MSIM3 for migration flows in the period 1985–90. Several observations can be made regarding this matrix. First, among the 756 (756 = 28 x 28 – 28) specific migration flows, there are 464 specific interprovincial flows that are over-predicted (those flows with negative signs), implying that about 61.4 percent of the total number of flows falls in the range of overestimation by MSIM3. The remaining flows (38.6 percent) in the matrix are underestimated by MSIM3. It is also noteworthy that variation in the underestimated migration flows (the cells with positive values) is smaller than those overestimated ones (the cells with negative signs).

Second, the distribution of the 464 overestimated migration flows varies among the 28 provinces viewed both as origins and destinations. Table 9 presents a summary of this distribution. When provinces are viewed as origins (or out-migration from origins), the number of overestimated flow ranges from eight flows in Shanxi to 25 flows in Guangxi. Only in three other origins is the number of overestimated flows greater than 20; and these are the provinces of Hebei (21), Inner Mongolia (21), and Hunan (22). When provinces are viewed as destinations (or in-migration to destinations), the number of overestimated flows varies between eight flows in Guangdong to 25 flows in Hunan. It is further found that in other four provinces the number of overestimated flows to destinations is found to be more than 20. These four provinces are Guizhou (24), Hebei (23), Inner Mongolia (21), and Shandong (21).

This pattern of overestimation is associated with underestimation in the remaining flows. In other words, examination of the residual patterns can be appropriately made in the context of the whole out- or in-migration fields (Evans and Pooler, 1987). For example, there are 25 overestimated out-migration flows from Guangxi, and only two flows are underestimated. These two flows are out-migrations to Zhejiang (with the relative residual of 0.23) and to Guangdong (with the relative residual of 0.48). It appears that the underprediction for these two flows causes all other out-flows from Guangxi to be over-predicted. The province of Guangdong accounts for about 72 percent of all out-migration from Guangxi. Guangdong is the destination that receives the largest amount of migrants from all other provinces, and Guangxi is one of largest origins for sending out-migrations (China, 1993). These two provinces are situated in close proximity to one another. The underprediction for out-migration flows from Guangxi to Guangdong reveals that the destination of Guangdong is strongly favored by migrants from Guangxi. On the other hand, in order to satisfy the origin and distance constraints, the underprediction for these two flows occurs at the expense of other flows being overestimated. Therefore, Guangdong, as the largest economically booming destination in China, exerts a considerable impact on the magnitude of other flows.

Table 8: Matrix of the relative residuals based on the observed and predicted inter-provincial migration flows, MSIM3, 1985–90.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
1	0	-0.65	-0.26	-0.09	-0.82	-0.86	-0.52	0.31	0.16	-0.01	0.48	0.04	0.29	-0.05	-0.04	0.01	0.34	-0.48	0.81	0.09	0.50	-0.64	0.34	0.18	-0.63	-0.11	-1.12	-0.15
2	-0.11	0	0.22	0.05	-0.43	-0.23	0.51	0.69	-0.05	-0.55	0.21	0.24	0.29	-0.60	-0.64	-0.31	0.01	-0.60	0.40	0.40	0.34	-1.35	0.16	-0.01	0.55	-0.16	0.16	-1.16
3	0.44	0.27	0	-0.34	-0.20	0.08	-0.14	0.09	-0.95	-0.72	-0.52	-1.04	**-1.97**	-0.99	-0.52	-0.68	-0.56	-0.17	-0.07	0.26	-0.09	**-2.91**	-0.42	-0.67	0.17	-0.18	-0.77	-0.28
4	0.21	0.12	-0.42	0	-0.06	-0.03	0.47	0.43	0.55	0.04	0.40	-0.21	0.07	0.18	-0.27	-0.01	-0.57	0.01	0.34	0.55	0.45	**-3.26**	0.40	0.22	0.53	0.23	0.13	0.05
5	-0.18	-0.25	-0.08	0.23	0	0.39	0.17	0.49	-0.17	-0.79	-0.45	-0.81	**-1.67**	0.16	-0.42	-1.21	-0.30	-1.17	-0.60	-1.01	-1.10	**-3.71**	**-1.55**	-0.27	-0.17	**-2.04**	-0.66	0.11
6	0.00	-0.23	0.02	-0.44	0.40	0	0.05	0.19	-0.73	-0.45	-0.11	0.05	-0.40	-0.40	-0.10	0.13	-0.10	-1.53	0.23	0.44	-0.07	-0.51	-0.04	-0.17	0.61	0.74	0.46	-1.03
7	0.00	-0.39	-0.07	0.03	0.31	-0.03	0	0.13	-0.32	-0.60	-0.50	-0.07	-0.12	0.12	0.19	-0.37	-0.72	-0.61	-0.14	0.33	-0.64	-0.35	-0.41	-0.42	0.65	0.24	**-1.67**	-0.19
8	-0.06	0.16	0.23	-0.91	0.49	0.12	-0.62	0	-0.45	-0.34	-0.45	0.10	-1.06	-0.42	0.22	0.16	**-1.69**	-0.56	-0.12	0.55	-0.45	-1.19	0.27	-1.28	-0.78	**-1.63**	-0.55	**-1.78**
9	0.13	-0.57	-0.87	0.13	-0.14	-0.60	0.20	-0.85	0	-0.02	0.13	0.20	-0.06	-0.06	-0.16	-1.46	0.11	-0.48	0.60	0.05	0.30	-0.89	0.30	-1.11	**-3.42**	-0.48	-0.32	0.37
10	0.29	-0.28	-0.83	-1.40	**-2.97**	0.19	0.12	0.34	0.33	0	-0.23	-0.21	0.18	-0.26	-0.44	-0.11	-0.62	**-1.97**	0.44	-0.06	-0.56	-0.96	0.00	-0.21	0.33	0.43	-0.22	0.30
11	0.25	0.06	-0.24	0.30	-0.16	0.47	-0.60	0.38	-0.37	-0.51	0	-0.34	0.22	0.18	-0.17	0.11	-0.09	-0.60	0.53	-0.01	-0.36	0.41	0.10	0.39	0.60	0.70	0.58	-0.02
12	0.43	0.21	-1.78	-0.16	**-1.56**	0.47	0.06	0.20	0.19	0.16	-0.21	0	0.29	0.05	-0.15	-0.54	-0.53	-0.36	-0.40	0.03	-0.80	-2.12	-0.24	-0.58	-0.31	0.10	0.21	0.44
13	0.14	-0.92	**-1.72**	0.52	**-1.69**	-0.24	-0.29	0.66	-1.34	-0.61	0.22	-0.35	0	0.24	-0.05	-0.46	-0.01	-0.83	0.55	-0.04	-0.05	-1.69	-0.63	0.53	0.53	0.21	-0.47	-0.67
14	-0.34	0.42	**-2.27**	**-1.71**	-0.38	-0.81	0.07	-0.01	-0.27	0.08	0.28	-0.14	0.39	0	-0.12	0.15	-0.39	-0.86	0.45	-0.45	-1.11	-0.59	-0.04	**-1.59**	0.33	-0.42	**-2.53**	0.35
15	0.16	0.05	-1.04	0.06	-0.31	0.28	0.29	0.46	-0.52	-0.16	0.19	-0.26	-0.60	-0.55	0	-0.04	-0.14	-0.56	-0.07	0.47	-0.52	-2.78	-0.14	-0.54	0.44	-0.12	-1.04	0.21
16	0.45	0.12	-0.41	-0.17	-0.64	0.03	-0.22	0.34	-0.18	-0.22	-0.30	-0.40	-0.87	-0.50	-0.31	0	0.02	-0.53	0.41	-0.90	-0.56	-1.15	-0.50	-0.10	0.22	0.57	0.12	0.57
17	0.22	0.28	-0.45	-0.14	**-2.04**	0.31	0.18	0.67	-0.25	-0.23	0.39	-0.14	0.11	-0.40	**-4.50**	0.40	0	0.06	0.60	0.22	0.18	-0.01	0.30	0.20	0.30	-0.97	**-2.24**	0.28
18	-0.56	-0.53	**-10.9**	0.48	-0.46	**-1.69**	-0.77	0.01	-0.13	-0.28	-0.13	-0.29	-0.18	-0.06	-1.27	-0.12	-0.45	0	0.68	0.56	-0.72	-0.17	0.37	-1.10	-0.92	-0.53	**-1.90**	-0.02
19	0.28	-0.77	-1.38	-0.38	**-1.85**	-0.27	0.03	-0.37	0.30	-0.29	0.28	-0.93	0.19	0.20	-0.46	0.18	0.15	-0.44	0	0.30	-0.09	0.08	0.60	-0.12	0.62	-0.51	**-2.73**	-0.21
20	-1.25	-1.42	-0.58	*-1.55*	**-2.61**	**-2.98**	-0.80	**-2.69**	**-2.10**	-1.15	0.23	-1.17	-0.49	**-2.14**	**-1.52**	**-3.21**	-1.04	**-1.59**	0.48	0	**-2.21**	-1.45	-1.08	-0.81	**-5.14**	**-2.21**	**-9.55**	-0.95
21	-0.11	-0.42	-0.05	-0.09	-0.46	0.02	-0.05	-0.07	-0.45	0.04	0.05	-0.35	0.31	-0.72	-0.61	-0.36	-0.08	-0.51	0.50	-0.49	0	0.01	0.22	-0.31	-0.92	0.03	-0.63	0.33
22	-0.77	-0.32	0.20	-0.01	-0.89	-1.37	0.10	**-1.51**	-0.82	0.40	0.30	0.20	0.38	-0.29	0.09	0.48	-0.22	-0.15	0.35	-0.75	-0.35	0	-0.16	-1.26	**-2.57**	**-4.94**	**-5.25**	**-2.91**
23	-0.87	-0.54	-0.51	-0.46	-1.09	**-1.63**	0.41	**-1.55**	-0.52	0.20	0.41	0.27	0.02	-0.09	0.51	-0.10	-0.22	-0.17	0.24	-1.39	0.11	-0.40	0	-1.18	0.04	**-2.36**	**-4.44**	-0.85
24	-0.04	-0.92	-0.18	0.07	0.17	0.18	0.29	0.10	-0.03	-0.15	-0.24	0.08	-0.26	-0.25	-0.14	0.19	-0.69	-0.37	0.26	0.26	0.03	-0.58	0.47	0	0.35	0.56	-0.82	0.46
25	-0.28	-2.30	-0.34	0.33	0.27	-0.10	0.11	-0.78	-0.28	-0.17	-0.01	-0.64	0.33	-0.08	-0.54	0.03	-1.09	-0.44	0.02	-3.92	-0.37	-3.75	-0.80	-0.03	0	0.55	0.34	0.65
26	-0.40	-0.65	-0.06	-6.22	-0.79	-0.23	-0.62	-0.50	0.19	0.11	0.51	0.34	0.05	-0.12	0.23	0.34	0.50	-0.07	-0.04	0.41	0.08	-0.33	-2.53	0.11	-0.03	0	-0.31	0.58
27	0.03	**-3.94**	-0.01	-1.16	0.23	-0.60	0.42	-1.38	0.11	0.12	-0.65	0.51	-0.35	-0.49	-0.49	0.54	-0.19	-1.04	0.50	**-1.60**	-0.65	-6.31	-0.93	-0.13	0.33	0.69	0	0.68
28	-0.07	-0.12	-0.36	-1.30	-0.93	-0.93	0.18	-1.31	0.50	-0.05	0.07	0.04	-0.04	-1.25	0.16	0.27	0.10	-0.34	-0.82	-0.41	-0.04	-1.50	-0.52	0.07	0.20	-0.78	-0.55	0

Notes: (1) The numbers of the first column and first row stand for the provincial ID number as the order specified in Table 7.
(2) Large residuals are defined as those equal to or beyond -1.50, and such residuals are bolded.

Table 9: Number of overestimated (negative sign) flows for province as both origin and destination, based on the residual matrix of Table 8, 1985–90.

Province	Out-migration flow	In-migration flow	Province	Out-migration flow	In-migration flow
Beijing	15	13	Shandong	17	21
Tianjin	13	18	Henan	17	14
Hebei	21	23	Hubei	11	20
Shanxi	8	17	Hunan	22	25
In. Mongolia	21	21	Guangdong	15	8
Liaoning	15	16	Guangxi	25	12
Jilin	18	10	Sichuan	18	19
Heilongjiang	18	11	Guizhou	18	24
Shanghai	16	19	Yunnan	18	15
Jiangsu	16	19	Shaanxi	13	20
Zhejiang	12	12	Gansu	18	10
Anhui	14	16	Qinghai	15	15
Fujian	18	13	Ningxia	16	20
Jiangxi	18	20	Xinjiang	18	13

Third, large overestimated flows, which are defined as the relative residuals beyond or equal to –1.50 (bolded in the residual matrix), are found to be mainly the in-migration flows to the five provinces of Hebei, Inner Mongolia, Guizhou, Qinghai, and Ningxia. It can be observed and calculated from Table 8 that 65 specific flows are under this category, and these five provinces account for 33 such flows. Guizhou and Ningxia each have nine such overestimated flows.

Finally, the largest overestimated flow is found in out-migration from Hunan to Hebei, and the relative residual is -10.9. This prediction error results mainly from the migrant stock variable for this particular flow. Specifically, migration from Hunan to Hebei in 1982–87 accounted for 49.4 percent in the total out-migration from Hunan, whereas this figure was found to be only 1.83 percent in 1985–90. In fact, for the origin of Hunan province, the correlation coefficient between out-migration in 1982–87 (used as the migrant stock) and out-migration in 1985–90 is only 0.2106. Hence, this largest overestimated flow is caused mainly by asymmetry between the migrant stock and out-migration in 1985–90. The asymmetry has to do with the return migration from Hunan to Hebei, which occurred in the early 1980s. Between the mid-1960s and 1970s, coal mining in Hunan was assisted with a large number of coal workers dispatched from Hebei, and they returned to their origin in the early 1980s (Shen and Tong, 1992).

MSIM3 includes two additional variables, the investment and migrant stock. The variable of migrant stock exhibits more importance in explaining the migration flows. The large overestimated flows can be explained by the asymmetry between the migrant stock and the migration in the period 1985–90. The relative residual in the out-migration flow from Hunan to Hebei is an obvious example already shown above. Although other large relative residuals may be involved in contextual circumstances different from the case of Hunan-to-Hebei migration, they mainly are caused by such asymmetry.

CONCLUSION

The purpose of this paper is to show an analysis and discussion of the results of the calibration of the conventional spatial interaction model and the models with the additional variable(s). The empirical verification of the models employed the two interprovincial migration data sets of China. These include the 1982–87 and 1985–90 migration flow data, consisting of two 28 x 28 data matrices. Constraints were applied to the models, and these are the origin and cost constraints. The constraints may be also considered the calibration criteria. The cost constraint employed in the present study is the matching of the observed and predicted logarithmic mean migration distance, which ensures that optimum parameters for the models are achieved.

The results of the calibration show that all of the models with the additional variable(s) are capable of distributing migration flows with a much-improved degree of accuracy, in comparison with the conventional model with inclusion of the distance variable only. The calibration has therefore provided empirical support for the validity and utility of the multivariate approach to the spatial interaction modeling of migration. However, the results obtained on the basis of this approach do not necessarily imply that just any additional variables included into the model would result in a corresponding improvement in accuracy in the prediction of migration flows. This is evident in the performance levels between MSIMs 2 and 3.

The errors in prediction of interprovincial migration were examined by utilizing the relative residual measure. The results of analyzing the residuals show that the errors in prediction of migration flows by the MSIM3 were reduced significantly, in comparison with the CM. This is in accordance with the results indicated in the analysis of the goodness-of-fit statistic. It is also indicated from the relative residual matrix that Guangdong, one of the largest booming provinces, exerted a considerable influence on China's interprovincial migration flows. The large overestimated flows (with the relative residual beyond –1.50) illustrate mainly the historical flows that were not followed closely by subsequent population migrations. Overall, it can be concluded that the multivariate approach to the spatial interaction modeling of migration pursued in the present study is a valid one and it would be useful to apply it to the analysis of migration systems other than China's.

ACKNOWLEDGMENTS

We are indebted to Professor Avinoam Meir for his encouragement, and to two anonymous referees for their constructive comments. The opinions expressed in this paper are those of the authors and not necessarily of the institutions with which they are affiliated.

NOTE

1. We have made experimental calibration using reverse flows (i.e., s_{ji} of 1982–87) to replace s_{ij} in MSIM2 and MSIM3. The results show that the percentage of migrants misallocated is 21.63 percent and 20.77 percent, respectively, suggesting the performance is worse than those using migrant stock s_{ij} flows, as compared with the results in Table 3. This meant that using the reverse flows to estimate migration is not appropriate in this particular case.

REFERENCES

Alonso, W. (1968) Predicting best with imperfect data. *Journal of the American Institute of Planners,* 34(4):248–255.

Banister, J. (1987) *China's Changing Population.* Stanford, California: Stanford University Press.

Batty, M and March, L. (1976) The method of residues in urban modelling. *Environment and Planning A,* 8:189–214.

Bernstein, T.P. (1977) Urban youth in the countryside: Problems of adaptation and remedies. *The China Quarterly,* 69:75–108.

Brown, L.A. (1996) Development Aspects of Migration in Third World settings: A simulation with implications for urbanization. In Geyer, H.S. and Kuntuky, T.M. (eds.) *Differential Urbanization: Integrating Spatial Models.* London: Arnold, 264–290.

Cai, Fang (1999) Spatial patterns of migration under China's reform period. *Asian and Pacific Migration Journal* 8(3):313–327.

Cheng, T. and Selden, M. (1994) The origins and social consequences of China's Hukou system. *The China Quarterly,* 139:644–668.

China, Population Census Office (1993) *Tabulation of Population Census of the People's Republic of China.* Beijing: China Statistical Publishing House (in Chinese).

China, State Statistical Bureau (1988) *Tabulations of China One Percent Population Sample Survey, National Volume.* Beijing: Department of Population Statistics, State Statistical Bureau (in Chinese).

——. (1996) *China Regional Economy: A Profile of 17 Years of Reform and Opening-up.* Beijing: China Statistical Publishing House (in Chinese).

Chisholm, M. (1996) Letters to the editor: How useful are measures of economic potential? *Environment and Planning A,* 28:1895–1900.

Evans, N.J. and Pooler, J. (1987) Distance deterrence effects in constrained spatial interaction models of inter-provincial migration. *Canadian Journal of Regional Science*, 10:259–279.

Fan, C.C. (1996) Economic opportunities and internal migration: A case study of Guangdong province, China. *The Professional Geographer,* 48(1):28–45.

Fan, C.C. and Huang, Y. (1998). Waves of rural brides: Female marriage migration in China. *Annals of the Association of American Geographers,* 88(2):227–251.

Fotheringham, A.S.(1997) Trends in quantitative methods 1: Stressing the local. *Progress in Human Geography*, 21(1):88–96.

Fotheringham, A.S. and O'Kelly, M.E. (1989) *Spatial Interaction Models: Formulations and Applications*. Dordrecht: Kluwer Academic Publishers.

Goldstein, S. (1990) China. In Nam, C.B., Serow, W.J. and Sly, D.F. (eds.) *International Handbook on Internal Migration*. New York, Greenwood Press, pp. 63–83.

Hay, A.M. (1991) Classics in human geography revisited: Commentary 2. *Progress in Human Geography*, 15(4):434–435.

Haynes, K.E. and Fotheringham, A.S. (1984) *Gravity and Interaction Models*. Beverly Hills, California: SAGE Publications, Inc.

Hugo, C.J. (1991) Migrant women in developing countries. In United Nations, *Internal Migration of Women in Developing Countries*. New York: United Nations, pp. 47–73.

Knudsen, D.C. and Fotheringham, A.S. (1986) Matrix comparison, goodness-of-fit, and spatial interaction modelling. *International Regional Science Review*, 10(2):127–147.

Kueh, Y.Y. (1992) Foreign investment and economic change in China. *The China Quarterly,* 131:637–690.

Kullback, S. (1959) *Information Theory and Statistics*. New York: Wiley.

Linge, G.J.R and Forbes, D.K. (1990) The Space Economy of China. In Linge, G.J.R and Forbes, D.K.(eds.) *China's Spatial Economy*. Hong Kong: Oxford University Press, pp. 11–34.

Ma, L.J.C and Wei, Y. (1997) Determinants of state investment in China, 1953–1990. *Tijdschrift voor Economische en Sociale Geografie,* 88(3):21–225.

Ma, L.J.C. and Xiang, B. (1998) Native place, migration and the emergence of peasant enclaves in Beijing. *The China Quartely,* 155:546–581.

Massey, D.S. (1986). The social organisation of Mexican migration to the United States. *The Annals of the American Academy of Political and Social Science*, 487:102–113.

——. (1990). Social structure, household strategies, and the cumulative causation of migration. *Population Index,* 56(1):3–26.

Massey, D.S., Arango, J., Hugo, G., Kouaouci, A., Pellegrino, A., and Taylor, J.D. (1993) Theories of international migration: A review and appraisal. *Population and Development Review,* 19(3):431–466.

Montgomery, J.D. (1991) Social networks and labor-market outcomes: Toward an economic analysis. *American Economic Review*, 81(5):1408–1418.

Naughton, B. (1988) The third front: Defense industrialization in the Chinese interior. *The China Quarterly*, 115:351–386.

Ottensmann, J.R. (1997) Partially constrained gravity models for predicting spatial interactions with elastic demand. *Environment and Planning A*, 29(6):975–988.

Plane, D.A. (1981) Estimation of place-to-place migration flows from net migration totals: A minimum information approach. *International Regional Science Review*, 6(1):33–51.

——. (1982) An information theoretic approach to the estimation of migration flows. *Journal of Regional Science*, 22(4):441–456.

Pooler, J. (1985) Derivation of a generalized family of spatial interaction models: An information theoretic approach. *Modelling and Simulation*, 16:199–207.

——. (1994) An extended family of spatial interaction models. *Progress in Human Geography*, 18(1):17–39.

——. (1995) Modelling spatial interaction without distances: The use of prior spatial flows. *Geographical Systems*, 2:309–324.

Raiser, M. (1998) Subsidising inequality: Economic reforms, fiscal transfers and convergence across Chinese provinces. *The Journal of Development Studies*, 34(3): 1–26.

Ravenstein, E.G. (1885). The law of migration. *Journal of the Royal Statistical Society*, 48:167–227.

Shannon, C.E. and Weaver, W. (1949) *The Mathematical Theory of Communication*. Urbana: University of Illinois Press.

Shen, J. (1999) Modelling regional migration in China. *Environment and Planning A*, 31:1223–1238.

Shen, Y.M and Tong, C.Z. (1992) *China's Population Migration*. Beijing: State Statistical Publishing House (in Chinese).

Snickars, F. and Weibull, J.W. (1977) A minimum information principle: Theory and practice. *Regional Science and Urban Economics*, 7:137–168.

Thomas, E.N. (1968) Maps of Residuals from Regression. In Berry, B.J.L. and Marble, D.F. (eds.) *Spatial Analysis: A Reader in Statistical Geography*. New Jersey: Prentice Hall.

Thomas, R.W. and Hugget, R.J. (1980) *Modelling in Geography: A Mathematical Approach*. New Jersey: Barnes and Noble Books.

Webber, M.J. (1979) *Information Theory and Urban Spatial Structure*. London: Croom Helm.

White III, L.T. (1979) The road to Urumchi: Approved institutions in search of attainable goals during pre-1968 rustication from Shanghai. *The China Quarterly*, 79:481–510.

Wilson, A.G. (1967) A statistical theory of spatial distribution models. *Transportation Research*, 1:253–269.

——. (1971) A family of spatial interaction models and associated developments. *Environment and Planning A*, 3:1–32.

——. (1974) *Urban and Regional Models in Geography and Planning*. London: John Wiley & Sons.

Wilson, S.R. (1976) Statistical notes on the evaluation of calibrated gravity models. *Transportation Research*, 10:343–345.

Yabuki, S. (1995) *China's New Political Economy: The Giant Awakes*. Boulder: Westview Press.

Zhang, Q. (1988) The formation of China's migration policy. *Chinese Population Science*, 2:35–38 (in Chinese).

Zhang, X., Wu, Z. and Chen, L. (1997) Age difference among the rural labour force in interregional migration. *Chinese Journal of Population Science*, 9(3):193–201.