Shen J, 1999, Modelling regional migration in China: estimation and decomposition, *Environment and Planning A*, 31, 1223-1238

++

Modelling regional migration in China: estimation and decomposition

Jianfa Shen

Department of Geography and the Center for Environmental Studies

The Chinese University of Hong Kong, Shatin, N.T., Hong Kong.

email: jianfa@cuhk.edu.hk

Abstract. This paper considers the issue of identifying the effects of spatial structure and the origin and destination attributes on interregional migration. A decomposition approach is developed based on migration models. The inter-provincial migration data in China over the period 1985-1990 are used to estimate a gravity migration model, an extended gravity model, a Poisson gravity model and a multilevel Poisson model which are then used to decompose the various effects on migration in China.

1 Introduction

Two kinds of approach have been used for modelling migration. The first approach uses age, gender, origin and destination-specific migration rates. This approach assumes that the number of migrants is determined by the size of population and the rate of migration and has been adopted in multiregional population models (Rees and Wilson, 1977; Rogers, 1995; Shen, 1994; 1996; Shen and Spence, 1996; 1997). Recent research (van Imhoff, van der Gaag, van Wissen and Rees, 1995) considered methods to reduce the number of population parameters to be estimated in the population system. Plane (1993) considered the problems of using fixed interregional transition probabilities. He explored models that utilized density-dampened and destination-population weighted transition probabilities to overcome the problem.

The second approach focused on modelling migration flows directly, explicitly using distance, origin and destination populations to explain migration. The classical gravity model only considers three variables (Hua and Porell, 1979). Extended demo-economic models often consider other socio-economic and environmental factors (Isserman, 1985; Stillwell and Congdon, 1991). Such migration models have rarely been linked with multiregional population models due to difficulties in calibration. First, migration flows need to be modelled by detailed age-specific groups. Second, many explanatory variables need to be projected first to project migration. Third and more importantly, the mechanism of how distance and spatial structure affect spatial migration flows is not well understood. Finally, it is well known that different model specification and different estimation methods may produce different parameter estimates and goodness of fit (Congdon, 1991). However, no further study has been made to show how different calibration of the model will affect the estimated contribution of the spatial structure and the origin and destination attributes on migration.

Indeed, many studies on migration have focused on the issue of model specification. A lengthy debate in the 1970s (Cliff *et al*., 1974, 1976; Curry *et al*., 1975; Johnston, 1973, 1975) and Sheppard (1979) focused primarily upon the problems of autocorrelation in the mass terms and the map pattern effect in the distance decay parameters. It has been shown that origin-specific distance decay terms are likely to be related to the map pattern (Fotheringham, 1981; Johnston, 1973, 1975). Fotheringham (1984; 1991) argued that destinations may be related by forces of agglomeration or competition.

This paper is concerned with identifying the effects of various factors including spatial structure, origin and destination attributes and random effects on migration. Total in-flows and out-flows will be decomposed into several components reflecting the effects of various factors. The inter-provincial migration data from the 1990 census in China will be used. The paper is organized as follows. A brief introduction to inter-provincial migration in China is provided in section 2. In section 3, an approach to decompose migration flows based on migration models will be developed. Section 4 specifies and estimates four migration models for China. Section 5 examines the decomposition results. Some conclusions are reached in section 6.

2 Inter-provincial migration in China 1985-1990

Since the late 1970s, the volume of internal migration in China has been increasing due to more relaxed migration policies. The direction of internal migration has also undergone a transition from east to west migration in the pre-reform period to west to east migration in the reform period (Shen, 1995; 1996).

The exact size of migration is not clear until the 1987 1% population sampling survey and the 1990 population census in China. Due to well-known inconsistency between household registration and the residence place of many individuals, the 1990 census considered following people as migrants over the five year period July 1, 1985 - July 1, 1990: those whose their official household registrations had changed and those who had left their place of registration for over a year.

No migration data on migrants moving into Xizan (Tibet) were collected. Thus migration between Xizan and other regions in China will not be considered here.

The migration data set used here refer to migrations among 29 provincial regions in mainland China except Xizan between 1985 and 1990. Figure 1 shows the provincial regions in China. Here, migration is measured by the number of migrants using a transition approach instead of a movement approach.

It may be of concern that the inter-provincial migration data among 29 areas in China may be inadequate to fully describe the migration process in China. However, this is the only spatial level that migration data for China as a whole are available and similar data have been widely used in migration studies on China. The conventional population weighted distance measure has been criticized for its implicit bias as more migrants are moving over shorter distance. Boyle and Flowerdew (1997) proposed to use migration-weighted distance to overcome the problem. The method may be most effective for a spatial configuration when an area was circled by another area like inner and outer London. Whether it will be an effective measure in the case of China remains to be tested and is beyond the scope of this paper. On the other hand, this paper will use highway distances rather than distances calculated from simple centroids and this may make it more difficult to implement Boyle and Flowerdew's method.

Table 1 presents the in-migrations, out-migrations and net migrations to and from 29 regions. According to the 1990 census, there were 11 million migrants who moved between the 29 provincial regions over the five year period. Guangdong received a largest net gain of more than one million migrants due to its rapidly expanding economy and heavy inflow of capital from Hong Kong, Macao and foreign countries. Beijing, the capital of China, and Shanghai, the leading economic center of China, both received a net gain of more than half million migrants. The most populous province, Sichuan, is the largest loser of migrants during the period (0.87

million). The least developed coastal province, Guangxi, lost 0.45 million population through migration. It is interesting to note that the coastal province, Zhejiang, is the third largest net loser. The overall picture of migration gains in China is as follows. Those regions with rapidly expanding economies are usually the top destinations of migrants. The main source of migrants were not necessary the poorest regions in the country, but the medium and less developed regions. Relatively developed regions such as Zhejiang could also be a major source of migrants as skilled individuals may move to other places. The Zhejiang village in Beijing formed by migrants from Zhejiang is a well know example (Li, 1996). Migration efficiency is the ratio of net migration to the total of in and out migrations. Main origin and destination regions have high migration efficiency.

3 Decomposition of migration flows based on migration models

Researchers have recognized the important role of distance in the migration process. Distance is often used as an explanatory variable to describe the volume of migration, M_{ij}, between places i and j. But no attempt has been made to identify the contribution of distance and other variables to migration except in terms of the estimated parameters. Specifically, no answer has been attempted to such question as what percentage of out or in migrations from or to a region is due to its particular location or due to its own particular status. Answers to these questions may be useful to formulate scenarios for migration projection. Migration rates based approach for migration projection totally ignores the role of distance and other variables at origin and destination.

Assuming that a migration model has been properly specified and estimated, the expected migration flow, M^E_{ij}, between region i and j can be expressed by a migration model. There are various ways to specify the random component of a migration model which is related to model specification but not for migration projection. This section will focus on the expected migration flow and the specification of the random component will be discussed in the next section. Consider first the gravity model which can be expressed as follows using a distance variable, d_{ij}, and populations at origin and destination p_i and p_j respectively:

$$M_{ij}^{E} = a_0 p_i^{a_1} p_j^{a_2} d^{b}{}_{ij} \qquad (1)$$

Assuming that the above model represents the migration flows correctly, then the contributions of the origin and destination attributes and the spatial structure to total outflows and inflows might be estimated using the following approach.

The expected total outflows from region i can be expressed as follows:

$$M_{i}^{E} = \sum_{j,j\neq i} a_0 p_i^{a_1} p_j^{a_2} d^{b}{}_{ij} \qquad (2)$$

From the point of region i, the volume of total out-migrations is partly determined by its spatial configuration (distance) in relation to other regions. If other things are equal, the shorter the distances between region i and other regions, the larger the volume of the out-migrations from region i. We can estimate the following number of out-migrations from region i using average population size of the spatial system to remove the effect of different population size at origin and destination on migration:

$$M_{i}^{CS} = \sum_{j,j\neq i} a_0 \overline{p}^{a_1} \overline{p}^{a_2} d^{b}{}_{ij} \qquad (3)$$

It is proposed here that the average of M^{CS}_i ("CS" refers to Constant and Spatial structure indicating two components of M^{CS}_i), M^C, means the constant volume of out migrations if the spatial configuration is the same for all regions and it may be called the "constant effect", the remaining part of M^{CS}_i, M^S_i , after taking away the average reflects the effect of different spatial configuration on out migrations and it is named as the “spatial structure effect I: space”. Another part of the spatial structure effect which will be discussed next.

The “constant effect” and the “spatial structure effect I: space” can be calculated as follows:

$$M^C = \tfrac{1}{N}\sum_i \sum_{j,j\neq i} a_0 \overline{p}^{a_1} \overline{p}^{a_2} d^b{}_{ij} \qquad (4)$$

$$M_i^S = \sum_{j,j\neq i} a_0 \overline{p}^{a_1} \overline{p}^{a_2} d^b{}_{ij} - M^C \qquad (5)$$

Here N is the number of regions in the spatial system.

The migration flows from region i can also be affected by the distribution of destination populations in the spatial system. The migration from region i will be greater if a region with a larger population is very close to region i. This is part of the effect of spatial structure and may be called “spatial structure effect II: attribute distribution”. This is different from the “spatial structure effect I: space” which assumes that population is the same in all regions. It can be calculated using destination populations instead of average population in equation (5) and then taking away the effects identified previously:

$$M_i^D = \sum_{j,j\neq i} a_0 \overline{p}^{a_1} p_j^{a_2} d^b{}_{ij} - M_i^S - M^C \qquad (6)$$

The difference between the total of first three effects and the expected outflows from region i will reflect the effect of the attribute in region i called the “own attribute effect”. This is different from the effect of the distribution of attributes among other regions. It can be calculated by using population at region i instead of average population in equation (6) and then taking away the effects identified before:

$$M_i^O = \sum_{j,j\neq i} a_0 p_i^{a_1} p_j^{a_2} d^b{}_{ij} - M_i^D - M_i^S - M^C \qquad (7)$$

The first item in the right of equation is the expected total outflow from region i. It has now been decomposed into four components:

$$M_i^E = M^C + M_i^S + M_i^D + M_i^O \tag{8}$$

Similarly, total expected inflows to region j, M^E_{j}, can be calculated as follows and it can also be decomposed into four components.

$$M_j^E = \sum_{i,i\neq j} a_0 p_i^{a_1} p_j^{a_2} d^b{}_{ij} \tag{9}$$

The "constant effect" for each inflow is the same of the "constant effect" for each outflow. It represents the expected total inflows to a region if everything is equal and can be calculated as follows:

$$M^C = \tfrac{1}{N}\sum_j \sum_{i,i\neq j} a_0 \overline{p}^{a_1} \overline{p}^{a_2} d^b{}_{ij} \tag{10}$$

The "spatial structure effect I: space" for inflows to region j reflects the effect of the spatial location of origins on migrations into region j and can be calculated as follows:

$$M_j^S = \sum_{i,i\neq j} a_0 \overline{p}^{a_1} \overline{p}^{a_2} d^b{}_{ij} - M^C \tag{11}$$

The migration flows to region j is also affected by the distribution of origin populations in the spatial system. This is part of the effect of spatial structure and is called "spatial structure effect II: attribute distribution". It can be calculated as follows for inflows to region j :

$$M_j^D = \sum_{i,i\neq j} a_0 p_i^{a_1} \overline{p}^{a_2} d^b{}_{ij} - M_j^S - M^C \tag{12}$$

The difference between the total of first three effects and the expected inflows to region j reflect the effect of the attribute in region j called the "own attribute effect". It can be calculated as follows:

$$M_j^O = \sum_{i,i \neq j} a_0 p_i^{a_1} p_j^{a_2} d^b{}_{ij} - M_j^D - M_j^S - M^C \quad (13)$$

The expected total inflows to region j has now been decomposed into four components:

$$M_j^E = M^C + M_j^S + M_j^D + M_j^O \quad (14)$$

Consider now a general migration model which include several explanatory attributes, $X=[x_1,x_2,..x_m]$, at origin and destination regions as follows:

$$M_{ij}^E = a_0 v(X_i) w(X_j) f(d_{ij}) \quad (15)$$

The decomposition approach described above can be applied straightforwardly to the above general migration model by substituting $p^{a1}{}_i$, $p^{a2}{}_j$ and $d^b{}_{ij}$ with functions $v(X_i)$, $w(X_j)$ and $f(d_{ij})$ respectively. Average attributes need to be used where average population is used in the case of gravity model.

4 Estimating inter-regional migration models

Migration models need to be estimated first before the decomposition approach can be applied to identify various migration components. In this paper, four migration models will be estimated for China.

The first model is the gravity model in equation (1) and will be estimated using Least Squares Estimation (LSE) in log-linear form. The model is specified as follows:

$$\ln M_{ij} = \ln a_0 + a_1 \ln p_i + a_2 \ln p_j + b \ln d_{ij} + e_{ij} \quad (16)$$

Here e_{ij} is assumed to be a normal distributed random variable.

Indeed, migration between two areas is also affected by other environmental, social and economic factors. The second migration model used in this paper is an extended gravity migration model which also includes several socio-economic and demographic variables. The model is specified as follows:

$$\ln M_{ij} = \ln a_0 + \sum_k a_{1k} \ln x_{ik} + \sum_k a_{2k} \ln x_{jk} + b \ln d_{ij} + e_{ij} \qquad (17)$$

Eight variables are used in model (17) in addition to three variables in a gravity model. These variables are:

GNPIi	Annual GNP growth rate over the period 1981-1989 at origin
GNPIj	Annual GNP growth rate over the period 1981-1989 at destination
ILLIj	Percentage of illiterate and semi-illiterate population aged 15+ in 1990 at destination
AGRILi	Percentage of agricultural employment in total rural employment in 1990 at origin
AGRILj	Percentage of agricultural employment in total rural employment in 1990 at destination
POP8290i	Percentage of population increase between 1982 and 1990 census periods at origin
DENSITYi	Population density in 1990 at origin
DENSITYj	Population density in 1990 at destination

These data for provincial regions of China are available from DPS (1991) and SSB (1990, 1991 and 1993). These variables describe important demographic, socio-economic situations in various areas which may affect the migration process. They have been selected by stepwise regression.

It is noted that the above two migration models (16) and (17) are estimated to minimise the residual squares of logged number of migration which can be expressed as:

$$\textit{Minimize} \qquad \sum (\ln M_{ij} - \ln M_{ij}^{E})^{2} = \sum (\ln \tfrac{M_{ij}}{M_{ij}^{E}})^{2} \tag{18}$$

It is clear that the LSE estimation of log-linear migration model aims to minimise the sum of the squares of logged ratios of the real migration size to expected migration size. Thus a ratio of 4/3 will be treated as the same as 4000/3000 while in reality a large flow of 4000 is much more important than a minor flow of 4. This kind of estimation criteria often result in poor fitting of large migration flows (Congdon, 1991). Another problem of the LSE estimation is that the total numbers of the actual and expected migrants will be different due to log-transformation.

A Poisson model is a more realistic description of the migration process than a log-linear model (Flowerdew and Aitkin, 1982). The third migration model is a Poisson gravity model with three variables. The total of actual migrants will be equal to total expected migrants in the model.

$$M_{ij} = \exp(\ln a_0 + a_1 \ln p_i + a_2 \ln p_j + b \ln d_{ij}) + u_{ij} \tag{19}$$

Here, migration flow is assumed to be a Poisson distributed variable and u_{ij} is the random residual.

In recent years, multilevel models have been developed to model the random variation at different group or regional levels (Jones, 1991). Boyle and Shen (1997) used a multilevel modelling approach to explore the relationship between migration and individual level and regional level factors. In terms of spatial migration, some origin or destination specific processes might be in operation which will affect origin and destination-specific migrations systematically. Thus a second level based on origin or destination can be specified for multilevel migration modelling. The fourth migration model is a multilevel Poisson migration model which also includes several socio-economic and demographic variables. The model is specified as follows:

$$M_{ij} = \exp(\ln a_0 + \sum_k a_{1k} \ln x_{ik} + \sum_k a_{2k} \ln x_{jk} + b \ln d_{ij} + e_{0i}) + u_{ij} \quad (20)$$

Here, migration flow is assumed to be a Poisson distributed variable at level one and u_{ij} is the random residual at level one. The level two is defined on the basis of origin regions. This is based on many empirical findings that migration from origins is more stable depending mainly on demographic factors than migration to destinations depending on both economic and other factors (Shen, 1996). It is assumed that there is random variation at level two which is represented by a normally distributed random variable e_{0i}. The model can be estimated using Mln software and Macros with a LOG link (Yang, Goldstein and Rasbash, 1996). Extra Poisson variation at level one is assumed as the model fails to converge if using fixed Poisson variation of one.

Table 2 presents the estimation results of the migration models (16), (17), (19) and (20). For the gravity model, all three variables are highly significant. In terms of logged number of migrants, the gravity model explained over 54% of its total variation. However, the model only explains 14.61% variation of the number of migrants for all flows.

In terms of logged number of migrants, the extended gravity model explained over 69% of its total variation which is much better than the gravity model. However, the model only explains 2.99% variation of the number of migrants for all flows and is much worse than the gravity model. The dramatic reduction in terms of model goodness of fit for both models is due the use of estimation objective equation (18) in log-linear regressions. It is clear that adding more variables to the gravity model does not necessarily improve the model performance in terms of the number of migrants.

In the Poisson gravity model, the population parameter became quite small. However, the model explained 23.23% of the total variation of migrants, better than two gravity models.

In the multilevel Poisson model, the variations at level one and level two are both significant. All variables except GNPIi (annual GNP growth rate over the period 1981-1989 at origin) and AGRILi (percentage of agricultural employment in total rural employment in 1990 at origin) are significant at 0.05 level. It is noted that the estimated parameters are different among various models reflecting the impact of model mis-specification. In the multilevel Poisson model, GNPIj (annual GNP growth rate over the period 1981-1989 at destination) has a positive value of 2.262 indicating the strong pulling effect of rapidly growing regions to migrants. On the other hand, ILLIj (percentage of illiterate and semi-illiterate population aged 15+ in 1990 at destination), AGRILj (percentage of agricultural employment in total rural employment in 1990 at destination) and DENSITYj (population density in 1990 at destination) have negative parameters of -0.7752, -1.9580, -0.3538 respectively. This indicates that, if everything is equal, regions with high percentage of illiterate and semi-illiterate population, high percentage of rural population engaged in agricultural employment and high population density are not attractive to migrants in China. It is clear that inter-provincial migration in China is been stimulated by the development, industrialisation and urbanisation processes. Two origin variables POP8290i (percentage of population increase between 1982 and 1990 census periods at origin) and DENSITYi (population density in 1990 at origin) also have negative parameters of -0.5871 and -0.3859 respectively. Thus, if everything is equal, regions with rapid population growth and high population density send out less migrants. It seems that the so called ‘PUSHING” mechanism is not operating effectively in the interregional migration. For inter-provincial migration, the origin population seems to be more important in determining the out-migration flow while in-migration is much more selective toward rapidly growing regions. Indeed, the parameter of origin population is much greater than that of destination population parameter.

Overall, the multilevel Poisson migration model explained over 39% of the variation of the number of migrants in all flows. This is a significant improvement over three other models. An alternative specification of a multilevel binomial model was also attempted but parameter estimates are generally close to the Poisson model. Nevertheless, over 60% of the migration variation remains unexplained indicating the complex nature of inter-provincial migration in China. Some descriptive analyses of internal migration in China tend to provide a convincing explanation of migration. But it is clear that such explanation may only provide a partial answer. Descriptive analysis is unable to predict how migration might change in response to changing regional systems. The decomposition of migration flows into several components might provide a way to capture such changes.

As the Poisson gravity model and multilevel Poisson model are better than other two models, they will be used in the decomposition of in and out migrations.

5 Decomposing regional outflows and inflows in China

The parameters estimated for the Poisson gravity model and multilevel Poisson model can now be used to decompose regional outflows and inflows in China. A random component might be defined as the difference between the real flow and estimated flow to indicate how well outflows and inflows have been modelled by a migration model.

Tables 3 and 4 presents the decomposing result of total outflows and inflows based on the Poisson gravity model. A relative random error is calculated as the percentage of random error out of expected flows. The estimated constant effect based on the Poisson gravity model is 402629. This is the number of in-migrations or out-migrations that each region is expected to receive or send if all regions have the same location in the spatial system and have the same values of all attributes. Other effects and the real and expected flows in tables 3 and 4 have been calculated as a percentage of the constant effect. The space effect is almost the same for in and out flows in each region and the effects of own attribute are in same direction. The effects of attribute distribution are relatively small.

The results from the multilevel Poisson model are presented in table 5 and 6. The estimated constant effect is 252819 for each region. This is much smaller than that of Poisson gravity model, but almost the same as gravity model. It is clear that the Poisson gravity model has used a larger constant effect to achieve a better overall performance than the gravity model. The space effect could be positive or negative depends on the particular location of a region in the spatial system. Some central located regions will send out extra migrants if other things remain equal. For example, the space effect on out-migration from Beijing is 57.13% while it is -68.59% for out-migration from Xinjiang.

The attribute distribution effects have different direction in two models. According to table 5, all regions have a positive effect of the spatial distribution of attributes on their out-migrations. This effect reflects whether the real distribution of a variable among destination regions will attract more out-migrations from a region than the case that each destination has the same average value of attributes. It looks like that each region send out extra migrants in an unbalanced spatial system.

The estimated space effects and own attribute effects based on two Poisson models are in the same direction. The own attribute effect represents the effect on out-migration if a region has large or small values of attributes than the average value in the spatial system. For the multilevel Poisson model, this effect is the combined result of the population size and other socio-economic variables. Such effects could be quite significant for some regions. For example, the own attribute effects on out-migration are 240.64% in Sichuan and -169.33% in Shanghai respectively. For some regions, the own attribute effect is not particularly larger than the space effect and the attribute distribution effect. For example, for out-migrations from Anhui, the space and attribute distribution effects are 42.22% (of the constant effect) and 95.53% respectively while the own attribute effect is 58.18%. This clearly indicates that it is important to consider the spatial structure in modelling interregional migration.

Finally, the random effect represents the component of migration which has not been explained by the model. This is closely related to the random error. The random errors of out-migrations in the multilevel Poisson model are smaller than those in the Poisson gravity model in most cases (tables 3 and 5). For example, the random error of out-migrations from Xinjiang is 229.13% in Poisson gravity model but reduced to 70.42% in the multilevel Poisson model.

It is interesting to examine the various effects on total in-migrations in various regions which is usually difficult to model. These effects are presented in table 6 and figures 2-5. According to figure 2, the space effect on in-migrations is the highest in the eastern coastal regions in China spanning from Zhejiang to Hebei.

Many regions in the western part of China have low accessibility in the spatial system and will not attract many migrants if other things are equal.

The spatial pattern of the attribute distribution effects on in-migrations in figure 3 looks unfamiliar at first. Due to the spatial distribution of population and other socio-economic variables, many regions in the central part of China have relative large positive effects. This means that a region will receive more migrants if its neighbouring regions have large populations or with poor socio-economic conditions. The own attribute effect represents the pulling force of the region to migrants due to their population size and/or socio-economic conditions. Several regions including Beijing, Heilongjiang, Jiangsu, Zhejiang, Guangdong have outstanding pull effects (table 6 and figure 4).

Finally, figure 5 presents the random effect on in-migration which is not explained by the multilevel Poisson model. A few regions including Liaoning, Anhui, Guangdong, Hainan, Sichuan and Xinjiang have large positive random effects and they are under predicted by the model. Table 6 also presents the random errors of in-migrations in various regions. Several regions still have a random error over 50%. This means that modelling interregional migration in China is not a easy task and further research is needed to explore new modelling techniques.

6 Conclusion

This paper considers the issue of identifying the effects of spatial structure and the origin and destination attributes on migration. A decomposition approach is developed based on migration models. A set of migration data in China are used to estimate various migration model which are then used to decompose the various effects on migration in China.

It is found that the multilevel Poisson migration model is the best model which explained over 39% of the variation. The multilevel Poisson model indicates that, if everything is equal, regions with high percentage of illiterate and semi-illiterate population, high percentage of rural population engaged in agricultural employment and high population density are not attractive to migrants in China. It seems that inter-provincial migration in China is been stimulated by the development, industrialisation and urbanisation processes. On the other hand, regions with rapid population growth and high population density send out less migrants. It seems that the so called ‘PUSHING” mechanism is not operating effectively in the interregional migration..

The space effect could be positive or negative depends on the particular location of a region in the spatial system. The space effects are almost identical to inflows and outflows. The space effect on in-migrations is the highest in the eastern coastal regions in China spanning from Zhejiang to Hebei.

Due to the spatial distribution of population and other socio-economic variables, many regions in the central part of China have relatively large positive attribute distribution effects. Several regions including Beijing, Heilongjiang, Jiangsu, Zhejiang, Guangdong have outstanding pull effects, the own attribute effect.

It is expected that various components of inflows and outflows of migrations can be used for formulating migration projection scenarios with adjustments. The mechanisms to achieve this and the ways to deal with more detailed age-groups is beyond the scope of this paper and will be the focus of further research.

Acknowledgement

This research is funded by a Direct Research Grant of the Chinese University of Hong Kong, Project code 2020328. Thanks are also due to Dr Paul Boyle and three anonymous referees for their useful comments.

References

Boyle P J and Flowerdew R, 1997, Improving distance estimates between areal units in migration models, *Geographical Analysis*, **29**(2), 93-107

Boyle P J and Shen J, 1997, Public housing and migration: a multi-level modelling approach, *International Journal of Population Geography*, **3**, 227-242

Cliff A D, Martin R L and Ord J K, 1974, Evaluating the friction of distance parameter in gravity models, *Regional Studies*, **8**, 281-286

Cliff A D, Martin R L and Ord J K, 1976, A reply to the final comment, *Regional Studies*, **10**, 341-342

Congdon P, 1991, An application of general linear modelling to migration in London and South East England, in: Stillwell J C H And Congdon P (eds), *Migration models: macro and micro approaches*. Belhaven, London. pp. 113-137

Curry L, Griffith DA and Sheppard E S, 1975, Those gravity parameters again, *Regional Studies*, **9**, 289-296

DPS, 1991, China Population Statistics Yearbook 1990, Dept. of Population Statistics, State Statistical Bureau. China Statistics Press, Beijing

Flowerdew R and Aitkin M, 1982, A method of fitting the gravity model based on the Poisson distribution, *Journal of Regional Science*, **22**, 191-202

Fotheringham A S, 1981, Spatial structure and distance-decay parameters, *Annals of the Association of American Geographers*, **71**, 425-436

Fotheringham A S, 1984, Spatial flows and spatial patterns, *Environment and Planning A*, **16**, 529-543

Fotheringham A S, 1991, Migration and spatial structure: the development of the competing destinations model, in J. Stillwell and P. Congdon (eds), *Migration Models: Macro and Micro Approaches*, Belhaven Press, London, pp. 57-72

Hua Chang-i and Porell F, 1979, A critical review of the development of the gravity model, *International Regional Science Review*, **4**(2), 97-126

Isserman A M, 1985, Economic-demographic modelling with endogenously determined birth and migration rates: theory and prospects, *Environment and Planning A*, 17, 25-45

Johnston R J, 1973, On frictions of distance and regression coefficients, *Area*, **5**, 187-191

Johnston R J, 1975, Map pattern and friction of distance parameters: a comment, *Regional Studies*, **9**, 281-283

Jones K, 1991, Specifying and estimating multi-level models for geographical research, *Transactions of Institute of British Geographers NS*, 16, 148-60

Li Cheng, 1996, Surplus rural laborers and internal migration in China: current status and future prospects, *Asia Survey*, XXXVI(11), 1122-1145

Sheppard E S, 1979, Gravity parameter estimation, *Geographical Analysis*, **11**, 120-132

Plane D A, 1993, Requiem for the fixed-transition-probability migrant, *Geographical Analysis*, **25**, 211-223

Rees P and Wilson A G, 1977, *Spatial Population Analysis*, Edward Arnold, London

Rogers A, 1995, *Multiregional Demography: Principles, Methods and*

Extension, Wiley, Chichester

Shen J, 1994, *Dynamic Analysis of Spatial Population Systems*, CATMOG 57, School of Environmental Sciences, University of East Anglia, Norwich

Shen J, 1995, Rural development and rural to urban migration in China 1978-1990, *Geoforum*, 26, 395-409

Shen J, 1996, Internal Migration and Regional Population Dynamics in China, *Progress in Planning*, 45, Part 3, 123-188

Shen J and Spence N A, 1996, Modelling urban-rural population growth in China, *Environment and Planning A*, 28, 1417-1444

Shen J and Spence N A, 1997, Modelling regional population growth in China, *Mathematical Population Studies*, **6**, 241-274

SSB, 1990, *Historical Statistics of Provinces in China 1949-1989*, State Statistical Bureau, China Statistics Press, Beijing

SSB, 1991, *China Statistics Yearbook 1991*, State Statistical Bureau, China Statistics Press, Beijing

SSB, 1993, *Tabulations on the 1990 Population Census of the People's Republic of China*, State Statistical Bureau, China Statistics Press, Beijing

Stillwell J C H And Congdon P (eds), 1991, *Migration models: macro and micro approaches*. Belhaven, London

van Imhoff E V, van der Gaag N, van Wissen L and Rees P H, 1995, Model selection for internal migration at the NUTS-2 level, *Technical Report*, Netherlands Interdisciplinary Demographic Institute, The Hague

Yang M, Goldstein H and Rasbash J, 1996, *Mln Macros for Advanced Multilevel Modelling*, Multilevel Models Project, Institute of Education, University of London

Table 1 Inter-provincial migrations in China 1985-1990

Region	In-migrations	Out-migrations	Net migration	Migration efficiency
Guangdong	1257218	250494	1006724	0.668
Beijing	671671	132148	539523	0.671
Shanghai	664756	132562	532194	0.667
Liaoning	540735	294996	245739	0.294
Tianjin	244065	72194	171871	0.543
Jiangsu	789555	620478	169077	0.120
Shanxi	306578	218472	88106	0.168
Hubei	429914	346274	83640	0.108
Shandong	607446	534842	72604	0.064
Xinjiang	341573	277412	64161	0.104
Hainan	150059	105977	44082	0.172
Ningxia	91883	56609	35274	0.238
Qinghai	115087	102141	12946	0.060
Fujian	250962	238387	12575	0.026
Yunnan	249462	277432	-27970	-0.053
Nei Mongol	254264	303129	-48865	-0.088
Shaanxi	309690	362349	-52659	-0.078
Jiangxi	224412	293772	-69360	-0.134
Gansu	197175	280715	-83540	-0.175
Henan	474867	589626	-114759	-0.108
Jilin	237232	355532	-118300	-0.200
Guizhou	190056	312786	-122730	-0.244
Hebei	519147	645704	-126557	-0.109
Anhui	336665	533388	-196723	-0.226
Heilongjiang	367394	607485	-240091	-0.246
Hunan	271036	527614	-256578	-0.321
Zhejiang	332311	632323	-300012	-0.311
Guangxi	142436	588889	-446453	-0.610
Sichuan	439130	1313049	-873919	-0.499

Table 2 Estimation result of gravity and Poisson migration models

Variables	Gravity model		Extended gravity model		Poisson model		Multilevel Poisson model	
	Para-meter	Standard error	Para-meter	Standard error	Para-meter	Standard error	Para-meter	Standard error
Constant (a_0)	-5.8744*	1.2883	6.0236*	1.4904	-1.0510	1.8000	2.4590	2.5530
Distance	-1.1000*	0.0593	-1.2431*	0.0534	-0.8623*	0.0589	-1.1040*	0.0408
Origin Population	0.7777*	0.0448	0.9063*	0.0537	0.6945*	0.0741	0.8318*	0.1364
Dest. Population	0.5356*	0.0448	0.7423*	0.0495	0.2720*	0.0631	0.6192*	0.0714
GNPIi			0.5372*	0.2294			0.1906	0.5013
GNPIj			1.2504*	0.2296			2.2620*	0.2406
ILLIj			-0.4233*	0.1009			-0.7752*	0.1116
AGRILi			-0.8929*	0.2090			0.0927	0.5185
AGRILj			-2.2621*	0.2186			-1.9580*	0.2517
POP8290i			-0.8639*	0.1364			-0.5871*	0.2805
DENSITYi			-0.3821*	0.0429			-0.3859*	0.1024
DENSITYj			-0.2843*	0.0422			-0.3538*	0.0566
Level 1 variance					24170	1198	9753	494
Level 2 variance							0.0899	0.0322
R^2(logged M)	0.5425		0.6954					
R^2(unlogged M)	0.1461		0.0299		0.2323		0.3935	

Note: * significant parameter at 0.05 level

Table 3 Decomposition of total out-migrations in various regions based on Poisson gravity model estimation (%*)

Region	Real Flow	Expected Flow	Space	Attribute distribution	Own attribute	Random effect	Random error
Beijing	32.82	53.15	37.53	-8.26	-76.12	-20.32	-38.24
Tianjin	17.93	47.51	39.43	-5.88	-86.04	-29.58	-62.26
Hebei	160.37	180.83	41.41	-9.21	48.63	-20.46	-11.31
Shanxi	54.26	97.95	24.63	-3.80	-22.88	-43.69	-44.6
Nei Mongol	75.29	56.03	-8.73	-6.56	-28.68	19.26	34.37
Liaoning	73.27	85.51	-9.92	-5.41	0.83	-12.24	-14.32
Jilin	88.30	57.15	-18.25	-3.30	-21.29	31.15	54.51
Heilongjiang	150.88	64.56	-26.47	-4.34	-4.63	86.32	133.7
Shanghai	32.92	55.74	16.28	0.94	-61.47	-22.82	-40.93
Jiangsu	154.11	183.01	27.49	-2.09	57.61	-28.90	-15.79
Zhejiang	157.05	113.29	15.29	-6.87	4.87	43.76	38.63
Anhui	132.48	166.84	28.97	0.30	37.57	-34.37	-20.6
Fujian	59.21	64.76	-20.62	-1.89	-12.73	-5.55	-8.58
Jiangxi	72.96	109.16	12.59	-1.04	-2.40	-36.19	-33.16
Shandong	132.84	199.78	23.35	-6.67	83.09	-66.94	-33.51
Henan	146.44	208.50	24.83	-4.19	87.85	-62.05	-29.76
Hubei	86.00	146.67	18.20	-1.35	29.82	-60.66	-41.36
Hunan	131.04	143.27	7.40	-2.15	38.02	-12.22	-8.53
Guangdong	62.21	108.15	-19.06	-3.41	30.62	-45.94	-42.48
Guangxi	146.26	80.27	-22.85	-1.34	4.46	65.99	82.2
Hainan	26.32	19.76	-31.48	-0.47	-48.28	6.56	33.21
Sichuan	326.12	150.31	-20.45	-5.21	75.96	175.81	116.97
Guizhou	77.69	71.52	-17.75	-1.02	-9.71	6.16	8.62
Yunnan	68.91	71.39	-22.80	-3.24	-2.57	-2.48	-3.48
Shaanxi	90.00	91.39	6.56	-3.85	-11.33	-1.39	-1.52
Gansu	69.72	54.10	-11.20	-9.33	-25.36	15.62	28.86
Qinghai	25.37	17.04	-19.10	-4.18	-59.68	8.33	48.9
Ningxia	14.06	17.71	-18.13	-4.49	-59.67	-3.65	-20.61
Xinjiang	68.90	20.93	-57.13	-2.58	-19.36	47.97	229.13

Note: figures in this table are in percentages of the constant effect except that the random error is in percentage of expected flow

Table 4 Decomposition of total in-migrations in various regions based on Poisson gravity model estimation (%*)

Region	Real Flow	Expected Flow	Space	Attribute distribution	Own attribute	Random effect	Random error
Beijing	166.82	90.95	37.53	-8.70	-37.87	75.87	83.42
Tianjin	60.62	90.75	39.43	-3.39	-45.28	-30.13	-33.21
Hebei	128.94	146.64	40.00	-10.29	16.93	-17.70	-12.07
Shanxi	76.14	115.65	24.63	0.93	-9.91	-39.51	-34.16
Nei Mongol	63.15	72.50	-7.32	-7.44	-12.74	-9.35	-12.9
Liaoning	134.30	84.11	-9.92	-6.30	0.32	50.19	59.68
Jilin	58.92	69.54	-18.25	-3.03	-9.18	-10.62	-15.27
Heilongjiang	91.25	65.86	-26.47	-5.85	-1.81	25.39	38.55
Shanghai	165.10	93.28	16.28	8.53	-31.52	71.82	76.99
Jiangsu	196.10	151.26	27.49	2.96	20.82	44.84	29.64
Zhejiang	82.54	108.39	15.29	-8.75	1.85	-25.85	-23.85
Anhui	83.62	153.37	28.97	9.81	14.59	-69.75	-45.48
Fujian	62.33	74.76	-20.62	0.82	-5.44	-12.43	-16.63
Jiangxi	55.74	116.45	12.59	4.85	-0.99	-60.72	-52.14
Shandong	150.87	143.97	23.35	-6.72	27.34	6.90	4.79
Henan	117.94	152.60	24.83	-1.66	29.43	-34.66	-22.71
Hubei	106.78	134.59	18.20	4.93	11.46	-27.81	-20.67
Hunan	67.32	122.94	8.64	0.31	13.99	-55.62	-45.24
Guangdong	312.25	89.20	-19.06	-2.65	10.90	223.05	250.07
Guangxi	35.38	80.92	-22.85	1.97	1.79	-45.54	-56.28
Hainan	37.27	44.14	-31.48	3.13	-27.50	-6.87	-15.57
Sichuan	109.07	95.09	-20.45	-7.37	22.92	13.97	14.69
Guizhou	47.20	81.29	-17.75	3.19	-4.16	-34.08	-41.93
Yunnan	61.96	74.75	-22.80	-1.41	-1.04	-12.79	-17.11
Shaanxi	76.92	101.58	6.56	-0.23	-4.76	-24.66	-24.28
Gansu	48.97	66.05	-11.20	-12.02	-10.73	-17.08	-25.85
Qinghai	28.58	41.42	-20.34	-4.99	-33.25	-12.83	-30.98
Ningxia	22.82	43.05	-18.13	-5.17	-33.65	-20.23	-46.99
Xinjiang	84.84	31.17	-57.13	-2.59	-9.11	53.66	172.16

Note: figures in this table are in percentages of the constant effect except that the random error is in percentage of expected flow

Table 5 Decomposition of total out-migrations in various regions based on multilevel Poisson model estimation (%*)

Region	Real Flow	Expected Flow	Space	Attribute distribution	Own attribute	Random effect	Random error
Beijing	52.27	40.00	57.13	27.46	-144.58	12.27	30.67
Tianjin	28.56	46.03	58.99	79.56	-192.52	-17.47	-37.96
Hebei	255.40	275.89	56.85	49.14	69.90	-20.48	-7.42
Shanxi	86.41	159.10	30.63	41.09	-12.62	-72.69	-45.69
Nei Mongol	119.90	224.02	-14.50	20.97	117.56	-104.13	-46.48
Liaoning	116.68	142.29	-15.15	29.33	28.11	-25.60	-17.99
Jilin	140.63	163.70	-21.21	56.62	28.29	-23.08	-14.10
Heilongjiang	240.28	201.36	-31.00	18.58	113.78	38.93	19.33
Shanghai	52.43	37.88	27.74	79.47	-169.33	14.55	38.40
Jiangsu	245.42	220.31	41.39	18.81	60.12	25.11	11.40
Zhejiang	250.11	256.36	24.63	45.38	86.35	-6.25	-2.44
Anhui	210.98	295.92	42.22	95.53	58.18	-84.94	-28.71
Fujian	94.29	83.04	-29.56	30.76	-18.16	11.25	13.55
Jiangxi	116.20	170.32	12.07	41.30	16.95	-54.12	-31.78
Shandong	211.55	263.17	27.31	38.19	97.67	-51.62	-19.61
Henan	233.22	256.74	27.19	38.84	90.70	-23.52	-9.16
Hubei	136.97	213.51	18.04	33.95	61.52	-76.54	-35.85
Hunan	208.69	216.63	5.23	32.56	78.84	-7.94	-3.66
Guangdong	99.08	104.28	-27.31	7.65	23.94	-5.20	-4.99
Guangxi	232.93	115.00	-31.38	29.67	16.71	117.92	102.54
Hainan	41.92	23.22	-41.20	29.75	-65.33	18.70	80.56
Sichuan	519.36	320.25	-29.70	9.32	240.64	199.11	62.17
Guizhou	123.72	106.34	-25.17	22.20	9.31	17.38	16.35
Yunnan	109.74	131.36	-31.15	13.46	49.06	-21.63	-16.46
Shaanxi	143.32	145.38	2.25	25.71	17.42	-2.06	-1.42
Gansu	111.03	110.30	-15.01	2.62	22.70	0.73	0.66
Qinghai	40.40	61.69	-23.90	10.44	-24.85	-21.28	-34.51
Ningxia	22.39	21.67	-26.81	20.34	-71.87	0.72	3.34
Xinjiang	109.73	64.39	-68.59	5.87	27.10	45.34	70.42

Note: figures in this table are in percentages of the constant effect except that the random error is in percentage of expected flow

Table 6 Decomposition of total in-migrations in various regions based on multilevel Poisson model estimation (%*)

Region	Real Flow	Expected Flow	Space	Attribute distribution	Own attribute	Random effect	Random error
Beijing	265.67	277.61	57.13	4.50	115.98	-11.93	-4.30
Tianjin	96.54	100.64	58.99	10.48	-68.83	-4.10	-4.08
Hebei	205.34	210.39	55.02	4.20	51.17	-5.05	-2.40
Shanxi	121.26	194.34	30.63	30.40	33.32	-73.08	-37.60
Nei Mongol	100.57	153.23	-12.68	3.74	62.17	-52.66	-34.37
Liaoning	213.88	193.62	-15.15	15.80	92.97	20.26	10.47
Jilin	93.83	120.88	-21.21	35.92	6.18	-27.05	-22.38
Heilongjiang	145.32	213.30	-31.00	12.01	132.29	-67.98	-31.87
Shanghai	262.94	265.82	27.74	42.65	95.43	-2.88	-1.09
Jiangsu	312.30	386.48	41.39	20.62	224.47	-74.18	-19.19
Zhejiang	131.44	239.63	24.63	-12.56	127.57	-108.19	-45.15
Anhui	133.16	94.46	42.22	32.16	-79.92	38.71	40.98
Fujian	99.27	104.16	-29.56	14.90	18.82	-4.90	-4.70
Jiangxi	88.76	119.89	12.07	27.92	-20.11	-31.12	-25.96
Shandong	240.27	235.15	27.31	9.76	98.08	5.12	2.18
Henan	187.83	206.44	27.19	23.86	55.39	-18.61	-9.02
Hubei	170.05	162.47	18.04	29.20	15.24	7.58	4.66
Hunan	107.21	125.74	6.60	22.69	-3.55	-18.53	-14.74
Guangdong	497.28	356.27	-27.31	12.86	270.72	141.01	39.58
Guangxi	56.34	57.93	-31.38	21.13	-31.82	-1.59	-2.75
Hainan	59.35	36.54	-41.20	21.27	-43.52	22.81	62.42
Sichuan	173.69	133.09	-29.70	11.25	51.53	40.61	30.51
Guizhou	75.17	63.07	-25.17	31.14	-42.90	12.11	19.20
Yunnan	98.67	85.96	-31.15	21.88	-4.77	12.72	14.79
Shaanxi	122.49	123.13	2.25	29.49	-8.61	-0.64	-0.52
Gansu	77.99	70.12	-15.01	14.67	-29.54	7.87	11.23
Qinghai	45.52	29.58	-25.27	23.26	-68.41	15.94	53.91
Ningxia	36.34	26.98	-26.81	21.62	-67.83	9.36	34.69
Xinjiang	135.11	83.23	-68.59	7.27	44.55	51.87	62.32

Note: figures in this table are in percentages of the constant effect except that the random error is in percentage of expected flow

Figure 1. Provincial regions in China.

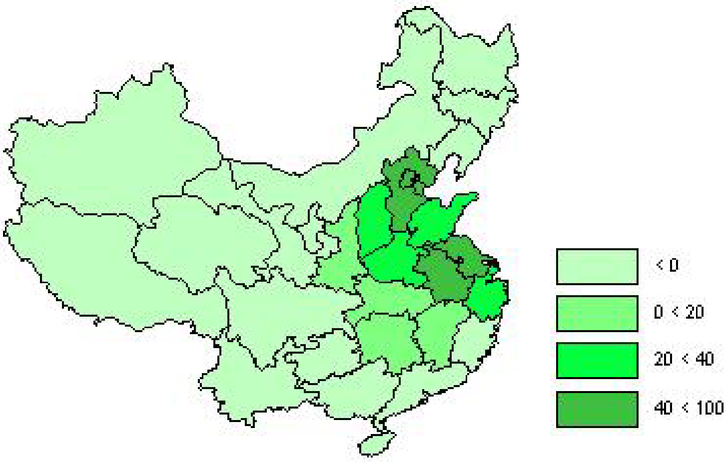

Figure 2 Space effect of in-migration based on Poisson model

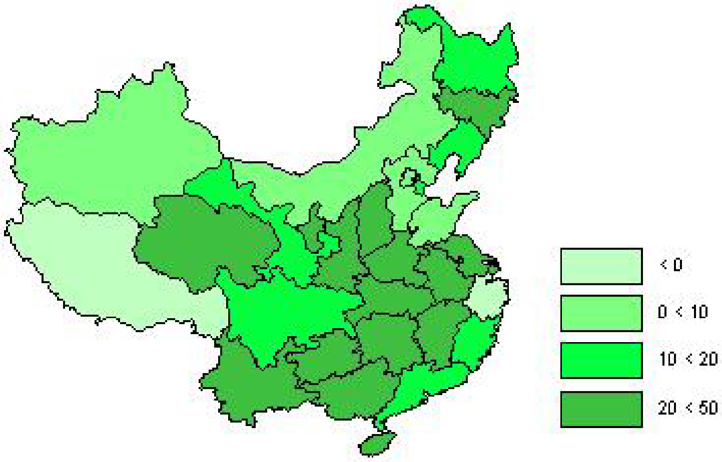

Figure 3 Attribute distribution effect of in-migration based on Poisson model

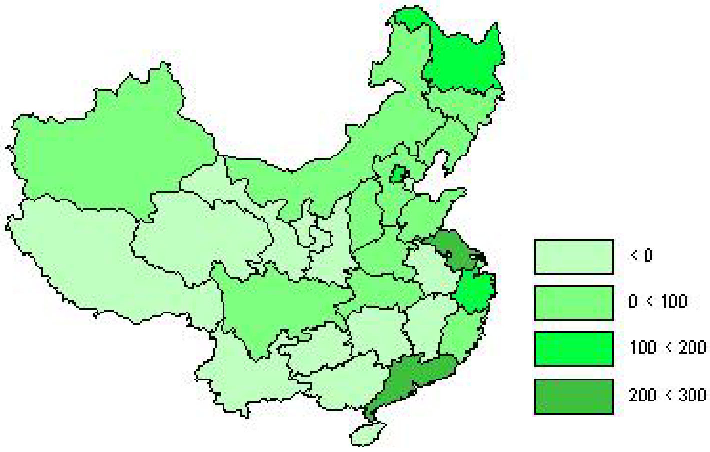

Figure 4 Own attribute effect of in-migration based on Poisson model

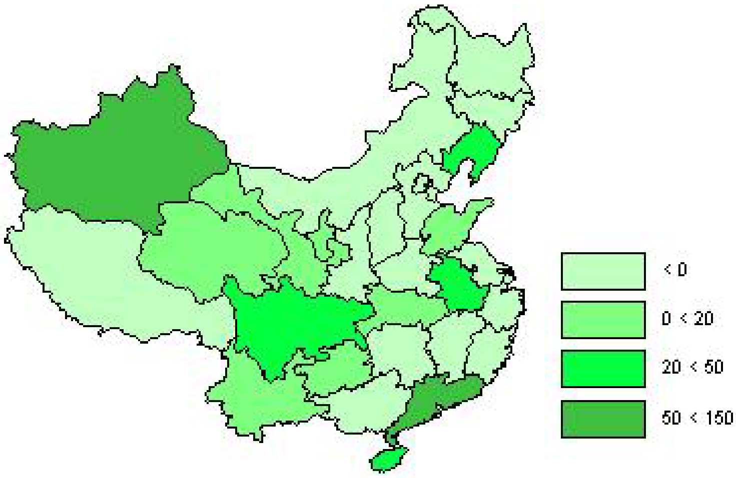

Figure 5 Random effect of in-migration based on Poisson model