한국지역지리학회지 제7권 제3호(2001) 107~120

A Spatial Statistical Approach to Migration Studies : Exploring the Spatial Heterogeneity in Place-Specific Distance Parameters

Sang-Il Lee*

This study is concerned with providing a reliable procedure of calibrating a set of place-specific distance parameters and with applying it to U.S. inter-State migration flows between 1985 and 1990. It attempts to conform to recent advances in quantitative geography that are characterized by an integration of ESDA(exploratory spatial data analysis) and local statistics. ESDA aims to detect the spatial clustering and heterogeneity by visualizing and exploring spatial patterns. A local statistic is defined as a statistically processed value given to each location as opposed to a global statistic that only captures an average trend across a whole study region. Whereas a global distance parameter estimates an averaged level of the friction of distance, place-specific distance parameters calibrate spatially varying effects of distance. It is presented that a Poisson regression with an adequately specified design matrix yields a set of either origin-or destination-specific distance parameters. A case study demonstrates that the proposed model is a reliable device of measuring a spatial dimension of migration, and that place-specific distance parameters are spatially heterogeneous as well as spatially clustered.

Key Words : place-specific distance parameters, migration, Poisson regression, ESDA, local statistics

1. Introduction

It has been one of the most fundamental geographical inquiries to provide reasonable explanations on 'flows' or 'movement' over space. The concept of spatial movement has been a pivotal conceptual construct in the sense not only that it allows geographers to formulate 'relative space' apart from 'absolute space' (Haynes and Fotheringham, 1984, 9), but also that it facilitates a *spatial* perspective on human behaviors by appropriating 'relation' with 'distance' (Gatrell, 1983), or by placing 'distance' over 'interaction' as its *explanan*. A collective set of quantitative approaches to spatial movements has been termed *spatial interaction models* (Fotheringham and O'Kelly, 1989). It is obvious that migration studies have played a pivotal role in the development and dissemination of the spatial interaction models.

* Ph.D. Candidate, the Department of Geography at the Ohio State University(slee@geography.ohio-state.edu)

Recent developments in quantitative geography have imposed tremendous impacts on spatial interaction models. The proliferation of spatial statistics and the advent and developments of GIS (Geographical Information Systems) as a general research platform have increasingly led to fundamental restructuring in quantitative geography (Fotheringham, et al., 2000). This new trend can be characterized by the integration between ESDA (Exploratory Spatial Data Analysis) and local statistics. ESDA is an extension of EDA (Exploratory Data Analysis) in general statistics and its ultimate objectives are to to describe and visualize spatial distributions, identify atypical locations or spatial outliers, discover patterns of spatial association, clusters or hot spots, and suggest spatial regimes or other forms of spatial heterogeneity (Anselin, 1994; 1998). This definition of ESDA effectively differentiates it from CSDA (Confirmatory Spatial Data Analysis) which characterizes much of the traditional quantitative geography and whose main goal is to *test* hypotheses, not to *formulate* them (Haining, 1990).

A local statistic is a statistically processed value assigned to each spatial unit in a whole study region. Unlike a global statistic that intrinsically captures an average trend for the entire region, a local statistic calibrates a *place-specific deviate* from the average trend. For example, a regression equation or a coefficient of determination derived from a regression is a global statistic, but another bi-product, a regression residual, is a local statistic.[1] In this sense, local statistics play a crucial role in the development of ESDA. When local statistics are visualized, a researcher is allowed not only to explore how a particular locale is deviated from a global trend, but also to identify spatial regimes that are spatial clusters of similar deviates (Anselin, 1995; 1998; 1999; Bao and Henry, 1996; Getis and Ord, 1996; Fotheringham, 1997; 2000; Fotheringham and Brunsdon, 1999; Unwin and Unwin, 1998). For example, albeit a positive correlation between two variables according to Pearsons correlation coefficient as a global statistic, some areas could show negative correlations and those areas further could be spatially clustered (Lee, 2001).

This new trend in quantitative geography provides spatial interactions models or more specifically migration studies with a fresh insight into the understanding of spatial movements. From primitive gravity models to more sophisticated entropy-maximizing models, most of the spatial interaction models are global in nature.[2] For example, a distance parameter calibrated by those models is nothing but an average trend and it does not offer any insight into *spatially varying effects of distance* in migration magnitude. Thus, the main objectives of this paper are to present the procedure of extracting place-specific distance parameters in a migration model and to demonstrate its usefulness in exploring the spatial heterogeneity of distance-decay effects with a real data set. Subsequently, I first present the Poisson regression as an alternative to gravity models and entropy-maximizing models. Secondly, a procedure of extracting place-specific distance parameters by means of the Poisson regression is presented. Thirdly, the usefulness and practicability of the procedure is discussed by applying it to an empirical data set of migration flows among 48 U.S. States between 1985 and 1990.

2. Calibration of Place-Specific Distance Parameters using a Poisson Regression

1) Limitations of gravity and entropy-maximizing models

A spatial interaction model is interested in models of the form :

$$Y_{ij} = \mu_{ij} + \varepsilon_{ij} \quad (1)$$

where Y_{ij} is an observed spatial interaction between i and j, μ_{ij} is an estimated spatial interaction, and ε_{ij} is an error term. Thus, spatial interaction models aim to develop a reliable way of calibrating μ_{ij}. A primitive form of the gravity model can be written as:

$$\mu_{ij} = k \cdot \frac{P_i^{\alpha} \cdot P_j^{\gamma}}{d_{ij}^{\beta}} \quad (2)$$

where P_i and P_j are total population in an origin i and a destination j, and d_{ij} is a distance between two locations. A usual way to estimate parameters (α, β, and γ) is to take a logarithm for the whole equation (2). Then we have :

$$\ln(Y_{ij}) = \ln(k) + \alpha\ln(P_i) + \gamma\ln(P_j) - \beta\ln(d_{ij}) + \hat{\varepsilon}_{ij} \quad (3)$$

Classical examples of applying equation (3) to migration studies can be found in Clayton (1977) and Flowerdew and Salt (1979). This procedure has a crucial drawback, let alone a number of conceptual and practical difficulties (see Senior, 1979). Obviously, equation (3) is a regression equation based on the Ordinary Least Squares (OLS) algorithm, which means that a log-normal distribution of Y_{ij} is required (Flowerdew and Aitkin, 1982). Such an assumption is not sustainable since spatial flows are discrete counts whose variance is very likely to be proportional to their mean value (Bailey and Gatrell, 1995, 353). As can be seen from the quadrat analysis (Thomas, 1977), it is more reasonable and statistically sounder to assume that counts of events occurring in areas follow the Poisson distribution. Second, there is no guarantee that flows predicted by equation (3) satisfy three constrains that are given respectively as :

(i) origin (production) constraint :

$$O_i = \sum_j \mu_{ij} = \sum_j y_{ij} \quad (4\text{-}1)$$

(ii) destination (attraction) constraint :

$$D_j = \sum_i \mu_{ij} = \sum_i y_{ij} \quad (4\text{-}2)$$

(iii) cost (impedance) constraint :

$$C = \sum_i \sum_j \mu_{ij} \cdot d_{ij} = \sum_i \sum_j y_{ij} \cdot d_{ij} \quad (4\text{-}3)$$

The origin and destination constraints require row-sums and column-sums in a predicted O-D matrix of migration flows to be identical to ones in an original O-D matrix, and the cost constraint is defined accordingly. Along with more sophisticated gravity models (see Haynes and Fotheringham, 1984), entropy-maximizing models, however, meet these requirements.[3)]

The entropy-maximizing model (Wilson, 1967; 1970) is specified by maximizing the *entropy function*, W, which is given by :

$$W(\{\mu_{ij}\}) = \frac{\left(\sum_i \sum_j \mu_{ij}\right)!}{\prod_{i,j} \mu_{ij}} \quad (5)$$

By maximizing equation (5) with requirements in (4) being conditioned, an optimal O-D matrix of flows is derived. Now, a flow between i and j locations is formally predicted by the equation :

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

where A_i and B_j are origin-related and destination-related balancing factors, and their values are estimated by the Lagrangian method. Although the entropy-maximizing model is conceptually sounder and practically more accurate than traditional gravity models, it has two crucial pitfalls. First, the model is a *predictive* device rather than a *explanatory* one, not only because it has been largely utilized in the context of planning, but because the specification itself tends to prevent itself from adjusting to embrace more exploratory variables (Flowerdew and Lovett, 1988). Second, it does not provide any statistical tests on estimated parameters, that is, whether they are statistically significant or not. This may be even more crucial when place-specific parameters are visualized to identify *significant* spatial clusters.

2) Poisson regression and a generalized spatial interaction model

With few exceptions (Flowerdew and Atkin, 1982; Flowerdew and Lovett, 1988; Scholten and van Wissen, 1985; Congdon, 1991; 1992; Flowerdew, 1991), Poisson regression models have rarely been utilized to specify migration flows. The problems discussed in connection with traditional gravity models and entropy-maximizing models can be solved by the Poisson regression or log-linear model. The Poisson regression is a special case of a class of GLM (Generalized Linear Model) that represents a synthesis including models such as linear regression, logit regression, binomial regression, and Poisson regression. When certain forms of population distributions in dependent variables allow for a linear transformation of a regression equation by way of a link function. When a linear transformation is done, a maximum likelihood algorithm yields a set of unbiased parameter estimators by maximizing a log-likelihood function (for more detailed introduction to GLM, see Gill, 2001).

In the context of migration flows, a series of derivatives for a log-likelihood function statistically guarantee those requirements presented in (4). Further, with the same set of variables, the entropy-maximizing model and the Poisson regression model yield an identical set of parameters (Tiefelsdorf and Boots, 1995). Without loss of generality, a general spatial interaction model based on the Poisson regression model is given by :

$$\ln(\mu_{ij}) = \hat{\mu} + \hat{\lambda}_i^O + \hat{\lambda}_j^D + \hat{\beta} d_{ij} + \sum_k \hat{\delta}_k X_{kij} \tag{7}$$

where $\hat{\mu}$ is an estimator for a base migration flow, $\hat{\lambda}_i^O$ and $\hat{\lambda}_j^D$ are parameters related respectively to the number of individuals at origin i and the number at destination j, X_{kij} denotes K other covariates such as population size at origin i or destination j, and $\hat{\delta}_k$ is a kth variables coefficient. The specification in (7) is called the *augmented doubly constrained spatial interaction* (Upton and Fingleton, 1989, 142).[4] Tiefelsdorf and Boots (1995) effectively

$$\begin{pmatrix} \ln(\mu_{11}) \\ \ln(\mu_{12}) \\ \ln(\mu_{13}) \\ \ln(\mu_{21}) \\ \ln(\mu_{22}) \\ \ln(\mu_{23}) \\ \ln(\mu_{31}) \\ \ln(\mu_{32}) \\ \ln(\mu_{33}) \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & d_{11} & p_1 \cdot p_1 \\ 1 & 1 & 0 & 0 & 1 & d_{12} & p_1 \cdot p_2 \\ 1 & 1 & 0 & 0 & 0 & d_{13} & p_1 \cdot p_3 \\ 1 & 0 & 1 & 1 & 0 & d_{21} & p_2 \cdot p_1 \\ 1 & 0 & 1 & 0 & 1 & d_{22} & p_2 \cdot p_2 \\ 1 & 0 & 1 & 0 & 0 & d_{23} & p_2 \cdot p_3 \\ 1 & 0 & 0 & 1 & 0 & d_{31} & p_3 \cdot p_1 \\ 1 & 0 & 0 & 0 & 1 & d_{32} & p_3 \cdot p_2 \\ 1 & 0 & 0 & 0 & 0 & d_{33} & p_3 \cdot p_3 \end{pmatrix} \cdot \begin{pmatrix} \hat{\lambda} \\ \hat{\lambda}_1^O \\ \hat{\lambda}_2^O \\ \hat{\lambda}_1^D \\ \hat{\lambda}_2^D \\ \hat{\beta} \\ \hat{\delta}_p \end{pmatrix} \tag{8}$$

demonstrated how a doubly constrained spatial interaction model with three areas is defined. By slightly modifying it, we have equation(8)

Here, the matrix consisting of independent variables usually called a design matrix Z should be noted. The row length of the matrix presents the number of all the cells in an O-D matrix. The first column of the matrix is composed of 1s that is required to fit a regression model. The second and third columns are dummy variables for the first and second areas as origins. It should be noted that zero is assigned to cells for the third area in those columns, since the area has been designated as a reference. The third and fourth columns are dummy variables for the first and second areas as destinations. Again, cells for the third area are given zero. The fifth column is a vector of distances among areas. The entries in the last column are products of origin population and destination population. It is necessary because it is impossible to put two columns simultaneously, one for original population and the other for destination population, due to a problem of collinearity.

In practice, a vectorized observed migration flows is set as a dependent variable and a design matrix as seen in (8) is set as a bundle of independent variables, a statistical package generates a vector of parameter estimators, one in the right side in (8). From this general specification of spatial interaction models, one may be able to define a particular specification for a particular research topic. In the next section, I will demonstrate how the general specification is modified to calibrate place-specific distance parameters.

3) Calibration of place-specific distance parameters

Although the notion of place-specific distance parameters have long been acknowledged, just a little literature has been dedicated to visualization and exploration of spatial patterns of the parameters (Fotheringham, 1981; Stillwell, 1991; Tiefelsdorf and Braun, 1997; 2000). Place-specific distance parameters can be divided into two categories: *origin-specific distance parameters and destination-specific distance parameters.* In the former, a large negative value for an area suggests that outflows to nearby areas are dominant from the area; in contrast, a small negative value indicates the areas larger out-migration field. In the latter, a larger negative parameter in an area suggests that most of inflows to the area are from nearby areas; in contrast, a small negative parameter indicates the areas larger spatial extension for attraction as a migration destination. An origin-specific distance parameters can be calibrated by the specification as :

$$\begin{pmatrix} \ln(\mu_{11}) \\ \ln(\mu_{12}) \\ \ln(\mu_{13}) \\ \ln(\mu_{21}) \\ \ln(\mu_{22}) \\ \ln(\mu_{23}) \\ \ln(\mu_{31}) \\ \ln(\mu_{32}) \\ \ln(\mu_{33}) \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & d_{11} & 0 & d_{11} & p_1 \\ 1 & 1 & 0 & d_{12} & 0 & d_{12} & p_2 \\ 1 & 1 & 0 & d_{13} & 0 & d_{13} & p_3 \\ 1 & 0 & 1 & 0 & d_{21} & d_{21} & p_1 \\ 1 & 0 & 1 & 0 & d_{22} & d_{22} & p_2 \\ 1 & 0 & 1 & 0 & d_{23} & d_{23} & p_3 \\ 1 & -1 & -1 & -d_{31} & -d_{31} & d_{31} & p_1 \\ 1 & -1 & -1 & -d_{32} & -d_{32} & d_{32} & p_2 \\ 1 & -1 & -1 & -d_{33} & -d_{33} & d_{33} & p_3 \end{pmatrix} \cdot \begin{pmatrix} \hat{\lambda} \\ \hat{\lambda}_1^O \\ \hat{\lambda}_2^O \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}^O \\ \hat{\delta}_p^D \end{pmatrix} \quad (9)$$

From (9), one may notice that cell values for the referenced area are given -1, not zero. This kind of dummy specification is called the *centered coding scheme* as opposed to the *cornered coding scheme*. The centered coding scheme is preferred mainly because all the parameters can be interpreted by reference to overall average trend, not to the referenced area. For example, $\hat{\beta}$ is an overall distance-decay parameter in the centered coding scheme, not the origin-specific distance-decay parameter for the third area such that $\hat{\beta}_1$ is the distance-decay parameter for the first area relative to the overall distance parameter, not relative to the third area. Thus, the actual origin-specific distance-decay parameters are computed by :

$$\hat{\beta}_i^O = \hat{\beta}^O + \hat{\beta}_i \tag{10}$$

Further, parameters for the referenced area which are not directly estimated should be computed by :

$$\hat{\lambda}_3^O = -\sum_i \hat{\lambda}_i^O \tag{11-1}$$

$$\hat{\beta}_3 = -\sum_i \hat{\beta}_i \tag{11-2}$$

A destination-specific distance-decay parameter can be calibrated by the specification as :

$$\begin{pmatrix} \ln(\mu_{11}) \\ \ln(\mu_{12}) \\ \ln(\mu_{13}) \\ \ln(\mu_{21}) \\ \ln(\mu_{22}) \\ \ln(\mu_{23}) \\ \ln(\mu_{31}) \\ \ln(\mu_{32}) \\ \ln(\mu_{33}) \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & d_{11} & 0 & d_{11} & p_1 \\ 1 & 0 & 1 & 0 & d_{12} & d_{12} & p_1 \\ 1 & -1 & -1 & -d_{13} & -d_{13} & d_{13} & p_1 \\ 1 & 1 & 0 & d_{21} & 0 & d_{21} & p_2 \\ 1 & 0 & 1 & 0 & d_{22} & d_{22} & p_2 \\ 1 & -1 & -1 & -d_{23} & -d_{23} & d_{23} & p_2 \\ 1 & 1 & 0 & d_{31} & 0 & d_{31} & p_3 \\ 1 & 0 & 1 & 0 & d_{32} & d_{32} & p_3 \\ 1 & -1 & -1 & -d_{33} & -d_{33} & d_{33} & p_3 \end{pmatrix} \cdot \begin{pmatrix} \hat{\lambda} \\ \hat{\lambda}_1^D \\ \hat{\lambda}_2^D \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}^D \\ \hat{\delta}_p^O \end{pmatrix} \tag{12}$$

When compared to (9), destination dummy and distance vectors replace origin ones, and a vector of origin population replace one of destination population. Following (10) and (11),

$$\hat{\beta}_j^D = \hat{\beta}^D + \hat{\beta}_j \tag{13-1}$$

$$\hat{\lambda}_3^D = -\sum_j \hat{\lambda}_j^D \tag{13-2}$$

$$\hat{\beta}_3 = -\sum_j \hat{\beta}_j \tag{13-3}$$

3. An Application to U.S. Migration Data

1) Data and spatial units

The data set consists of migration flows among 48 states (plus Washington D.C.) in the continental U.S. The data set has been obtained from the 1985 residency information registered by county in the 1990 U.S. Census.[5] This data set does not include immigration flows and the author eliminated migration flows from and to Alaska and Hawaii.

Figure 1 shows the spatial distribution of net migration. Since a pure intra-continental migration flows are concerned here, the pattern may not correspond to population growth pattern. Whereas the first two classes in the map indicate migration-lose, the last two classes suggest the prevalence of in-migration over out-migration. As commonly described, areas of migration-gain are concentrated on what has been called sun-belt regions and along both coastal lines, and areas of migration-lose are found in the traditional core regions including the Northeast and Midwest regions, and the Great Plain region. Figure 2-a and 2-b respectively show the spatial pattern of in- and out-migration. Those maps convey additional information about migration flows. For exampe, the low net-migration in the traditional industrial core region and the Texas region is associated with high in-migration but excessively high out-migration.

2) The spatial heterogeneity of place-specific distance parameters

The specification in (9) has been used to calibrate origin-specific distance parameters. A vector of $(\hat{\beta}_1, \Lambda, \hat{\beta}_n)$ has been mapped (Figure 3-a). Although actual distance parameters for States are computed by equation (10), the

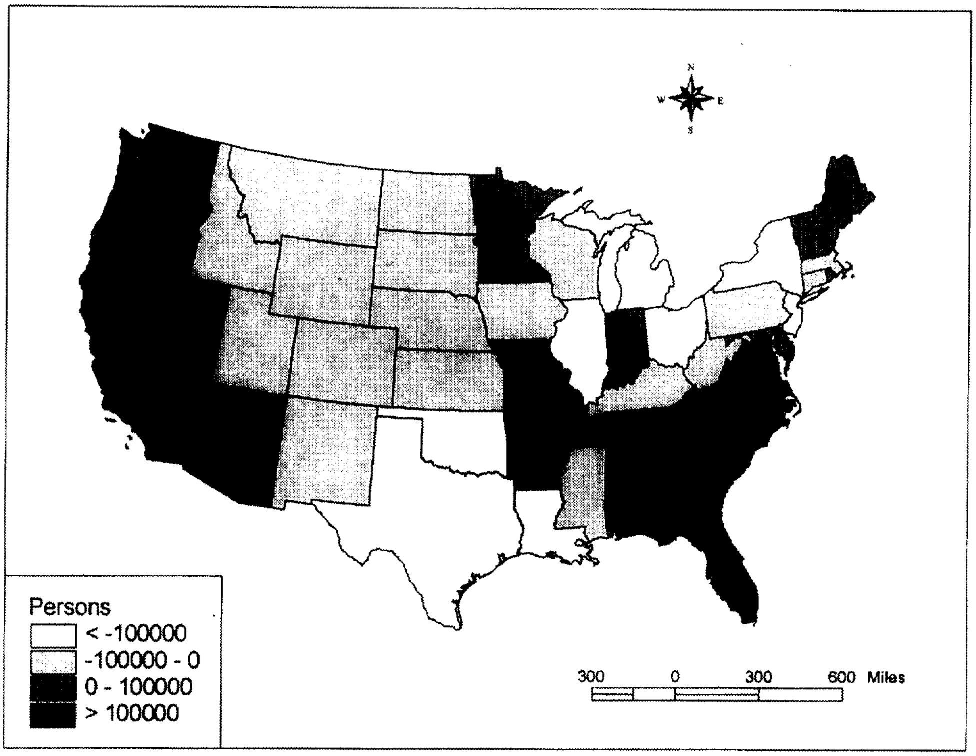

Figure 1. Net-Migration in U.S. States, 1985~1990

original parameters convey more intuitive information. If a locale follows the global trend, the parameter at the location will be zero. Further, positive values indicate that migrants from the area tend to take longer migration trips; in contrast, in areas with negative values, people move out under more pronounced distance-friction. The global distance parameter is calibrated at −0.140 (the original parameter was multiplied by 100), which means that, at each 100-mile increase, migration flows decrease by 13.06% when other conditions being held constant.[6)] If an areas distance parameter is 0.05 in Figure 2-a, out-migrants from the area experience the friction of distance only at a rate of 8.61% per 100 miles.

A set of destination-specific distance parameters can be calibrated by the specification in (12). Again, parameters originally estimated by a Poisson regression are preferred to actual ones calculated by equation (13-1). If an area has a positive parameter, it is implied that migrants into the area tend to take longer trips. In other words, those areas can be said to have larger migration fields, and thus may possess cities with higher ranks in a national urban hierarchy. On the contrary, migrants to areas with negative parameters tend to be originated from nearby states.

From Figure 3, one may recognize how much a single trend measure for distance-effects in migration distorts what has happened in reality. In short, distance parameters in migration phenomena are *spatially heterogeneous*. People in

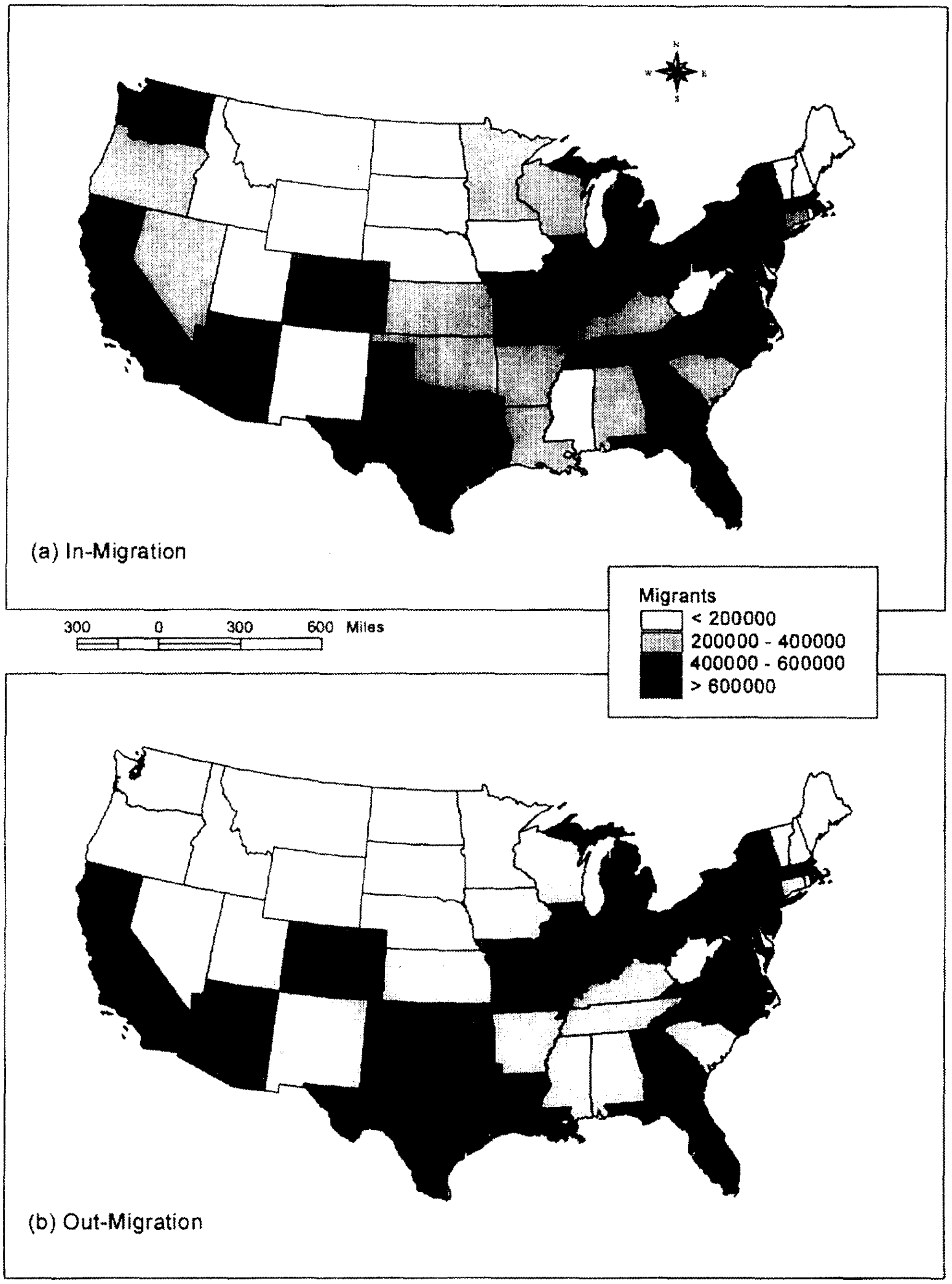

Figure 2. In-Migration and Out-Migration in U.S. States, 1985~1990

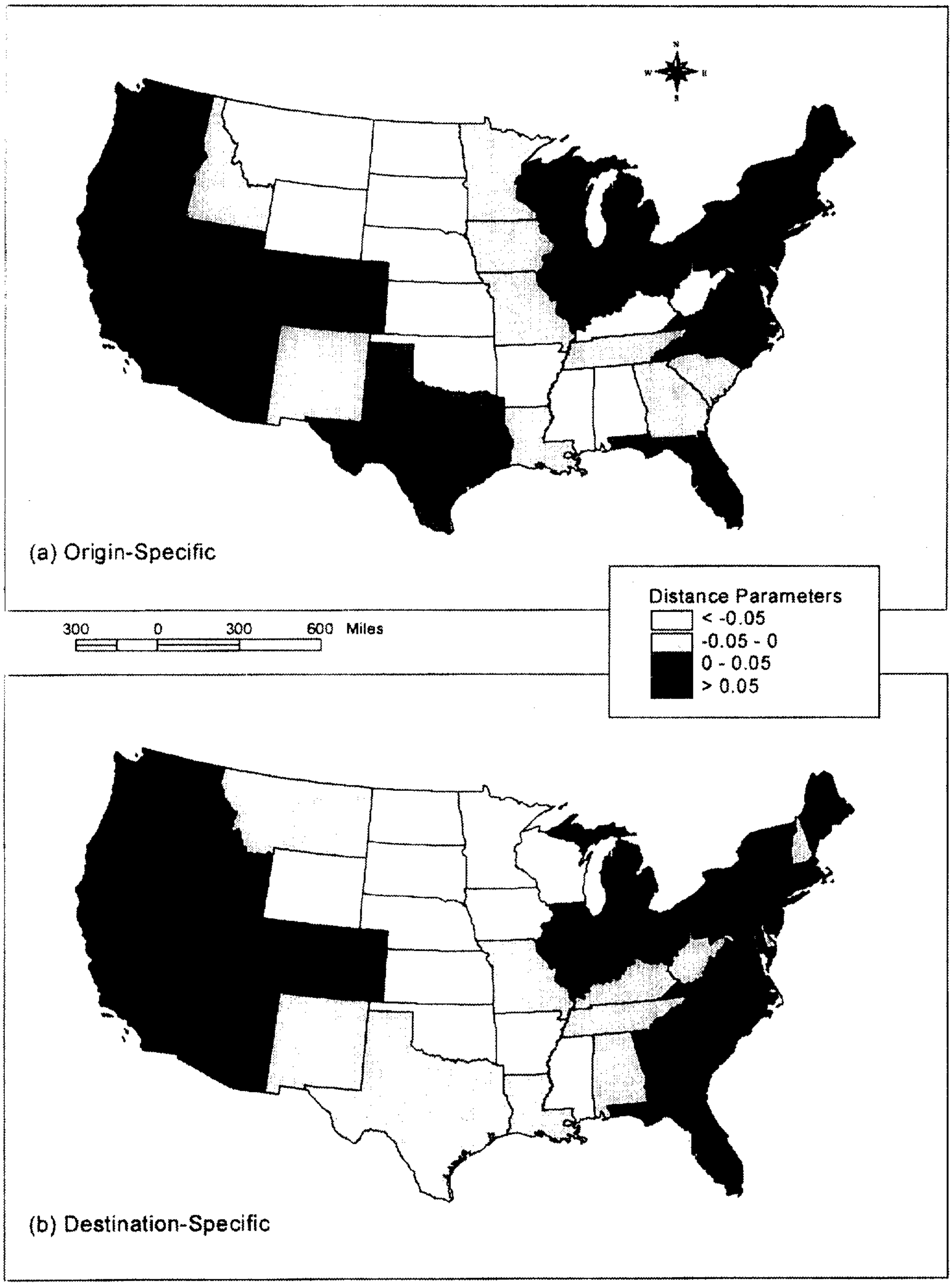

Figure 3. Place-Specific Distance Parameters in U.S. States

some areas have been more influenced by the friction of distance in their migration decisions than those in other areas. Conversely, some areas impose different level of the friction of distance on migrants from other areas in their decision on spatial movements to the area than other areas. In addition, distance parameters in migration are *spatially clustered*. Areas possessing locational similarities tend to show similar parameters. Out-migrants from a cluster of areas tend to behave similarly in their decision on spatial search; a cluster of areas tend to impose similar level of the friction of distance on people moving out of other areas.

Two spatial patterns in Figure 3 are very similar, which means that areas attracting longer -trip in-migrants tend to emit longer-trip out-migrants, and vice versa. This indicates that there is a strong tie between *spatial extents of attractivity and emissivity* in migration phenomenon. However, it should be noted that the spatial extent of attractivity does not necessarily relate to the number of in-migrations. Actually, the simple correlation between destination-specific distance parameters (Figure 3-b) and the number of in-migrants (Figure 2-a) is 0.47. For example, California, Florida, and Texas attracted more than a million migrants between 1985 and 1990 (see Figure 2-a). While both California and Florida show positive parameters, Texas is given a negative value, which indicates that in-migrants to Texas tend to be originated from nearby States. An opposite example is seen in Nevada and Oregon, where relatively small number of migrants to those States came from relatively distant parts of the U.S.

In accordance, the spatial emissivity is not necessarily associated with the number of out-migrants. A correlation coefficient between origin-specific distance parameters (Figure 3-a) and the number of out-migrants (Figure 2-b) is 0.46. For example, Colorado and Oklahoma emitted similar number of migrants (about 500,000), but they belong to opposite classes in terms of origin-specific distance parameters: 0.06 for Colorado and 0.06 for Oklahoma. This implies that, whereas out-migrants from Colorado tend to scatter throughout the country, out-migrants from Oklahoma tend to confine their destinations to nearby States.

One may observe that maps in Figure (3) are more spatially clustered than those in Figure (2). This means that areas with similar locational properties behave more correspondingly in *spatial extents* of migration than the amount of migration. Furthermore, spatial heterogeneity is also more pronounced in Figure (3) such that at least three distinctive spatial regimes are easily defined; say, the West, the Middle, and the East.

4. Conclusions

This paper aimed at providing a way of exploring the spatial heterogeneity in migration studies. A Poisson regression model was proposed to calibrate place-specific distance parameters in migration flows. The case study on U.S. inter-state migration flows between 1985 and 1990 demonstrates that the proposed model is a reliable device of measuring a spatial dimension of migration. In addition, map patterns with calibrated place-specific distance parameters evidenced that those parameters are spatially heterogeneous as well as spatially clustered.

As Fotheringham (2000) correctly point out, it is irony that, albeit a strong tradition of areal differentiation, quantitative geography has focused on spatial similarities rather than spatial differences, global generalities rather than local exceptions, and whole-map values rather than mappable statistics. Spatial interaction models, more specifically, migration models, should follow recent transitions occurring in overall geographic information sciences, that is, integration of ESDA (exploratory spatial data analysis) and local statistics. By decomposing intrinsically aspatial global statistics into place-specific or context-dependent local statistics, migration studies will gain more spatially saturated insights into migration phenomena.

Notes

1) It should be noted that one may need to make a clear distinction between a *spatial* local statistic and a *psudo-spatial* local statistic. The former is a value derived from a statistical computation with topological relationships among observations being taken into account, but the latter is a value simply assigned to an area. In this sense, a regression residual or a factor score is a psudo-spatial local statistics. Obviously, local spatial association measures belong to the category of spatial local statistics.
2) Earlier exceptions include Fotheringham (1981) and Ingram (1984).
3) A practical difference between more sophisticated gravity models and the entropy-maximizing model lies in how to estimate parameters. The former is based on an iterative procedure (Senior, 1979, Fig.3, 190) and the latter utilizes the Lagrangian multipliers. In general, the latter is superior to the former in terms of goodness-of-fit (Flowerdew and Lovett, 1988).
4) When the K-covariates are dropped from equation (7), the model becomes the doubly constrained model. When $\hat{\lambda}_i^O$ or $\hat{\lambda}_j^D$ is dropped, it becomes a destination-constrained (or attraction-constrained) spatial interaction model or an origin-constrained (or production-constrained) spatial interaction model.
5) More information about the data set can be found at http://www.census.gov/population/socdemo/migration/90mig.txt
6) The figure is computed by $100 \times \left(1 - e^{0.14}\right)$.

Acknowledgements

I wish to thank Michael Tiefelsdorf in the Department of Geography at the Ohio State University for his invaluable comments on materials presented here.

References

Anselin, L., 1994, Exploratory spatial data analysis and geographic information systems, Painho, M. (ed.), *New Tools for Spatial Analysis*, Eurostst, Luxembourg, 45~54.

Anselin, L., 1995, Local indicators of spatial association: LISA, *Geographical Analysis*, 27, 93~115.

Anselin, L., 1998, Exploratory spatial data analysis in a geocomputational environment, in Longley, P. A., Brooks, S. M., McDonell, R., and MacMillan, B. (eds.), *Geocomputation: A Primer*, John Wiley & Sons, Chechester, 77~94.

Anselin, L., 1999, Interactive techniques and exploratory spatial data analysis, Longley, P.A., Goodchild, M.F., Maguire, D.J., and Rhind, D.W. (eds.), *Geographical Information Systems, Vol. 1: Principles and Technical Issues*, 2nd Edition, John Wiley & Sons, New York, 253~266.

Bailey, T.C. and Gatrell, A.C., 1995, *Interactive Spatial Data Analysis*, Longman, New York.

Bao, S. and Henry, M.S., 1996, Heterogeneity issues in local measurements of spatial association, *Geographical Systems*, 3, 1~13.

Clayton, C., 1977, Interstate population migration process and structure in the United States, 1935 to 1970, *Professional Geographer*, 29, 177~181.

Congdon, P., 1992, Aspects of general linear modeling of migration, *The Statistician*, 41, 133

~153.

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.

Flowerdew, R. and Lovett, A., 1988, Fitting constrained Poisson regression models to interurban migration flows, *Geographical Analysis*, 20, 297~307.

Flowerdew, R. and Salt, J., 1979, Migration between labour market areas in Great Britain, 1970~1971, *Regional Studies*, 13, 211~231.

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

Fotheringham, A.S., 1997, Trends in quantitative methods I: stressing the local, *Progress in Human Geography*, 21, 88~96.

Fotheringham, A.S., 2000, Context-dependent spatial analysis: a role for GIS?, *Journal of Geographical Systems*, 2, 71~76.

Fotheringham, A.S. and Brunsdon, C., 1999, Local forms of spatial analysis, *Geographical Analysis*, 31, 340~358.

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

Fotheringham, A.S., Brunsdon, C. and Charlton, M., 2000, *Quantitative Geography: Perspectives on Spatial Data Analysis*, SAGE, Thousand Oaks.

Gatrell, A., 1983, *Distance and Space: A Geographical Perspective*, Oxford University Press, New York.

Getis, A. and Ord, J.K., 1996, Local spatial statistics: an overview, Longley, P. and Batty, M. (eds.), *Spatial Anlaysis: Modelling in a GIS Environment*, GeoInformation International, Cambridge, 261~277.

Gill, J., 2001, *Generalized Linear Models: A Unified Approach*, SAGE, Thousand Oaks.

Haining, R., 1990, *Spatial Data Analysis in the Social and Environmental Sciences*, Cambridge University Press, New York.

Haynes, K.E., 1984, *Gravity and Spatial Interaction Models*, SAGE, Beverly Hills.

Ingram, D.R., 1984, Historic evidence for variation in origin-specific distance-decay parameters, *Environment and Planning A*, 16, 123~125.

Lee S.-I., 2001, Developing a bivariate spatial association measure: an integration of Pearson's *r* and Moran's *I*, *Journal of Geographical Systems*, 3, in press

Scholten, H. and van Wissen, L., 1985, A comparison of the Loglinear interaction models with other spatial interaction models, Nijkamp, P., Leitner, H. and Wrigley, N. (eds.), *Measuring the Unmeasurable*, Martinus Nijhoff Publishers, Boston, 177~196.

Senior, M.L., 1979, From gravity modeling to entropy maximizing: a pedagogic guide, *Progress in Human Geography*, 3, 175~210.

Stillwell, J., 1991, Spatial interaction models and the propensity to migrate over distance, in Stillwell, J. and Congdon, P. (eds.), *Migration Models: Macro and Micro Approaches*, Belhaven Press, New York, 34~56.

Thomas, R., 1977, *An Introduction to Quadrat Analysis*, CATMOG series, No.12, Geo Books, Norwich.

Tiefelsdorf, M. and Boots, B., 1995, The specification of constrained interaction models using the SPSS loglinear procedure, *Geographical Systems*, 2, 21~38.

Tiefelsdorf, M. and Braun, G.O., 1997. The Migratory system of Berlin after the unification in the context of global restructuring, *Geographica Polonica*, 69 23~44.

Tiefelsdorf, M. and Braun, G.O., 2000, Inter-district migration patterns and trends within the city of Berlin, Germany, Paper presented at the Annual Meeting of the Association of American Geographers, Pittsburgh, PA, U.S.

Uniwn, A. and Unwin, D.J., 1998, Exploratory spatial data analysis with local statistics, *Journal of the Royal Statistical Society D: The Statistician*, 47, 415~421.

Upton, G.J.G. and Fingleton, B., 1989, *Spatial Data Analysis by Example: Volume 2 Categorical and Directional Data*, John Wiley & Sons, New York.

U.S. Census Bureau, http://www.census.gov/population/socdemo/migration/90mig.txt

Wilson, A.G., 1967, A statistical theory of spatial distribution models, *Transportation Research*, 1, 253~269.

Wilson, A.G., 1970, *Entropy in Urban and Regional Modelling*, Pion Limited, London.

인구이동 연구에 대한 공간통계학적 접근 : 장소-특수적 거리 패러미터의 추출과 공간적 패턴 분석

이 상 일*

국문요약

이 연구의 목적은 장소-특수적 거리 패러미터를 측정하는 방법론을 제시하고, 그것이 인구이동 연구에서 가지는 의미에 대해 미국의 48개 州間 인구이동자료를 사례로 검토해보는 것이다. 전통적인 인구이동 연구에서 추출하는 거리 패러미터는 인구이동량에 대해 거리가 가지는 평균적인 효과를 측정하는 것이다. 그러나, 그 평균적인 거리 패러미터는 모든 지역간 인구이동의 대표값일 뿐 인구이동에 있어 거리가 가지는 효과의 공간적 변이에 대해서는 아무런 통찰을 제공해 주지 못한다. 장소-특수적 거리 패러미터란 개개 소지역이 평균적인 거리 패러미터에 대해 가지는 상대적인 값이며, 거리가 인구이동에 대해 가지는 효과의 지역적 특이성을 측정하려고 한다.

이러한 연구는 최근 계량지리학 분야에서 발생하고 있는 변화에 부응하는 것이다. 1980년대 이후, 계량지리학은 공간통계학이라는 보다 폭넓은 개념의 확장과 일반연구환경으로서의 지리정보체계(GIS)의 성장으로 학문적 재구조화 과정 속에 있다. 이러한 재구조화 과정은 특정한 패러다임으로서의 탐구적 공간자료분석(ESDA)과 그것을 통계적으로 가능케 하는 국지 통계(local statistics)의 발달로 특징 지워진다. 통계적으로 가공되어 지역에 부여된 값으로 정의되는 국지 통계는 그것의 시각화를 효과적으로 수행하는 GIS와 결합함으로써, 시각화(visualization)와 과학활동으로서의 탐구(exploration)를 강조하는 탐구적 공간자료분석이라는 계량지리학의 새로운 패러다임을 효과적으로 수행하게 된다. 이러한 맥락에서, 장소-특수적 거리 패러미터는 하나의 국지 통계치로 인식될 수 있으며, 그것이 보여주는 공간적 패턴을 탐구하는 것은, 인구이동연구에서 탐구적 공간자료분석의 전형을 수행하는 것이라 볼 수 있다.

장소-특수적 거리 패러미터는 출발지-특수적 거리 패러미터와 도착지-특수적 거리 패러미터로 나뉘어 지는데, 이러한 패러미터를 추출하기 위해서는 특정한 통계기법이 요구된다. 이러한 패러미터를 추출하기 위해 전통적인 혹은 보다 진보된 형태의 중력모델이나 엔트로피-극대화 모델이 활용될 수 있지만, 본 논문은 포아송 회귀분석을 이용함으로써 패러미터의 추출이 가장 효과적으로 이루어짐을 논증하고 있다. 이 방법론은 1985년과 1990년 사이에 발생한 미국 48개 주간 인구이동량에 대한 사례연구에 적용되었다. 그 연구 결과는 장소-특수적 거리 패러미터의 공간성을 명확히 보여준다. 즉, 평균적 거리 패러미터로부터의 편기로 이해될 수 있는 장소-특수적 거리 패러미터들이 지역별로 상당한 차이를 보여줄 뿐만 아니라(공간적 이질성), 유사한 장소-특수적 거리 패러미터들이 공간적으로 집중되어 있음을 확인할 수 있었다(공간적 의존성).

지역차에 대한 강한 전통을 가지고 있는 지리학 내에서 태동한 계량지리학이 지역적 특이성을 무시하는 방향으로 발전해 온 것은 아이러니라 할 수 있다. 그것은 계량적 방법론의 한계라기 보다는 그 방법론을 사용하는 전통적 계량지리학자의 한계라고 보아야 할 것이다. 이러한 의미에서 본 연구는 최근 계량지리학의 경향을 인구이동연구에 적용한 사례임과 동시에 맥락 의존성을 강조하는 보다 폭넓은 과학운동의 계량지리적 반응이다.

주요어 : 장소-특수적 거리 패러미터, 인구이동, 포아송 회귀분석, 탐구적 공간자료분석, 국지 통계

(2001년 7월 29일 접수)

* 미국 오하이오 주립대 박사과정(slee@geography.ohio-state.edu)