LOCATION CHOICE MODEL WITH STRUCTURALIZED SPATIAL EFFECTS*

By Kazuaki MIYAMOTO**, Naoki SHIMOMURA***, and Varameth VICHIENSAN****

1. Introduction

It is well known that discrete choice model, especially logit model, has been continuously developed over the past ten years, mainly for the applications for transportation choice analysis. Those models are distinguished by specifying its deterministic and error components of the utility function to match the analysis purpose; resulting in different models such as multinomial logit model, nested logit model, mixed logit model, etc. However, discrete choice modeling for location analysis obtained less interest for research and development, which essentially requires the consideration of spatial effects among the choice set.

2. Spatial Effects in Discrete Choice Modeling

Although some of discrete choice models for transportation deal with spatial contexts, the location choice analysis is different from those in that it is essentially geographically referenced where choices are located continuously, making them mutually influenced in space.

2.1 Modeling Spatial Effects

In the spatial econometrical literature, spatial effects are often discussed on two aspects: spatial dependence and spatial heterogeneity (Anselin, 1988). Spatial dependence refers to situation where there exists similarity or dissimilarity among the neighborhood. If there is any systematic pattern in the spatial distribution of a variable, it is said to be spatially autocorrelated. If, for example, the pattern on a map is such that nearby or neighboring areas are more similar than more distant areas, the pattern is said to be positively spatially autocorrelated. The distribution of house prices or household income is usually positively spatially autocorrelated, since the wealthy households tends to live in exclusive neighborhoods, segregating themselves from lower income households. Spatial heterogeneity refers to situation where there exists non-uniformity of the effects of space. The spatial econometrical analysts usually know this as the situation where they obtain different model parameters if they use different sub-set of data for the model estimation. These spatial effects are inherent in both observable data and unobservable data used in modeling; however they are not treated properly in the location analysis. In other words, the spatial effects must be systematically accommodated in a discrete choice model for location analysis.

2.2 Location Choice Modeling with the Existence of Spatial Effects

Let us consider the residential choice of an arbitrary residential zone, which is locating adjacent to a zone which has immeasurable attractiveness. The traditional discrete choice model calculates the zonal utility from zone-specific explanatory variables such as residential convenience, etc. However, the zone has benefit in being close to the attractive zone. The benefit, however, becomes smaller if it is located farther away. This clearly exhibits spatial dependence. Therefore, spatial dependency must be incorporated by including the influence of the attractive zone into the utility of the residential zone. From decision maker viewpoint, the location choices are available in the continuous space, whose attractiveness is reduced with distance. This section has clearly shown that the traditional specification of discrete choice model cannot accommodate the situation when spatial effects are active.

Although there are existing studies trying to incorporate spatial effects in location choice modeling, they are not based on systematic specification of spatial effects. For example, Bhat *et.al.*(2003) presented a mixed logit model where the spatial dependence is modeled with an arbitrary spatial allocation rule, etc. The present paper presents a discrete choice model with a systematic specification of the spatial influences based on the knowledge of spatial econometrics. The rest of this paper is organized as follows. Section 3 presents the formulation of a new model for location analysis with structuralized spatial effects where the estimation technique is described. Section 4 presents an analysis with the simulated data while Section 5 presents a real case study. Finally, Section 6 concludes the paper with some remarks.

---

* Keywords: location analysis, spatial effects, mixed logit model
** Fellow of JSCE, D.Eng, Professor, Center for Northeast Asian Studies, Tohoku University
(Kawauchi, Aoba-ku, Sendai, Japan 980-8576 Tel. 022-217-7571, Fax. 022-217-7477)
*** Member of JSCE, M.Eng, Engineer, Penta-Ocean Construction Co., Ltd., Osaka Branch
(2-5-25 Umeda, Kita-ku, Osaka 530-0001Japan Tel. 06-6345-0631, Fax.06-6348-0347)
**** Member of JSCE, Ph.D., Research Fellow, Center for Northeast Asian Studies, Tohoku University
(Kawauchi, Aoba-ku, Sendai, Japan 980-8576 Tel. 022-217-7567, Fax. 022-217-7477)

3. A Discrete Choice Model with Structuralized Spatial Effects

The logit models (referred to as L model) are derived through the application of utility maximization concepts to a set of alternatives from which one, the alternative with maximum utility, is chosen. It is assumed that the utility of an alternative to a decision maker includes a deterministic term **V** and an additive error term **u**; that is,

$$\mathbf{U} = \boldsymbol{\beta}\mathbf{X} + \mathbf{u} = \mathbf{V} + \mathbf{u} \tag{1}$$

The deterministic term is commonly specified as linear in parameters $\boldsymbol{\beta}\mathbf{X}$, where the explanatory variables $\mathbf{X}$ represent the attributes of the alternatives. The well-known logit model is obtained by specifying that the error terms in (1) are independently and identically gumbel distributed across the alternatives. The spatial effects could be accommodated by specifying the utility function in various ways as follows.

3.1 Accommodating the Spatial Effect in the Deterministic Term

As discussed in Section 2 that there exist spatial interactions among the observable data in different zones. If the deterministic term in (1) is specified to be autocorrelated, we then obtain a logit model with autoregressive deterministic term (referred to as LAD model). The deterministic term is then expressed as

$$\mathbf{V} = \boldsymbol{\rho}_1\mathbf{W}_1\mathbf{V} + \boldsymbol{\beta}\mathbf{X} = (\mathbf{I} - \boldsymbol{\rho}_1\mathbf{W}_1)\boldsymbol{\beta}\mathbf{X} \tag{2}$$

where ρ_1 is the spatial interaction parameter, $\mathbf{W_1}$ is the spatial weight matrix, and **I** is the identity matrix. The specification is commonly used in Econometrics but rarely explored in discrete choice modeling. We may also specify the spatial dependence in the deterministic term differently such as directly specifying the correlation function, which results in the deterministic term $\mathbf{V} = \boldsymbol{\gamma}_1\mathbf{f}(\mathbf{d}) + \boldsymbol{\beta}\mathbf{X}$ where **f(d)** is a correlation vector as a function of distances between zones **d**.

3.2 Accommodating the Spatial Effect in the Error Term

In similar to the spatial interaction among the observable data, there is spatial effect among the unobservable data, which is known as the spatial autocorrelation. This may be accommodated by using the mixed logit model, which is characterized by the following error structure (McFadden *et.al.*, 2000)

$$\mathbf{u} = \boldsymbol{\xi} + \boldsymbol{\varepsilon} \tag{3}$$

where $\boldsymbol{\varepsilon}$ is the additive i.i.d. gumbel distributed variable vector. Ben-Akiva *et.al.*(1996) presented a mixed logit model (referred to as MLAE model) in which ξ in (3) is specified as follows

$$\boldsymbol{\xi} = \boldsymbol{\rho}_2\mathbf{W}_2\boldsymbol{\xi} + \boldsymbol{\mu} = (\mathbf{I} - \boldsymbol{\rho}_2\mathbf{W}_2)^{-1}\boldsymbol{\mu} \tag{4}$$

where ρ_2 is the spatial autocorrelation parameter, $\mathbf{W_2}$ is the spatial weight matrix, and μ is the standard-normal distributed variable vector. Similarly, the probit model with autoregressive error term could be specified with the utility function $\mathbf{u} = \boldsymbol{\rho}_2\mathbf{W}_2\mathbf{u} + \boldsymbol{\mu} = (\mathbf{I} - \boldsymbol{\rho}_2\mathbf{W}_2)^{-1}\boldsymbol{\mu}$ (McMillen, 1992).

3.3 A Mixed Logit Model with Structuralized Spatial Effects

This paper proposed a model that combines the concepts of the model in (2) and the model in (3) & (4) together, which results in a mixed logit model with autoregressive deterministic and error terms (referred to as MLADE model). The utility function of this model could be written as in (5).

$$\mathbf{U} = (\mathbf{I} - \boldsymbol{\rho}_1\mathbf{W}_1)^{-1}\boldsymbol{\beta}\mathbf{X} + (\mathbf{I} - \boldsymbol{\rho}_2\mathbf{W}_2)^{-1}\mathbf{u} + \boldsymbol{\varepsilon} \tag{5}$$

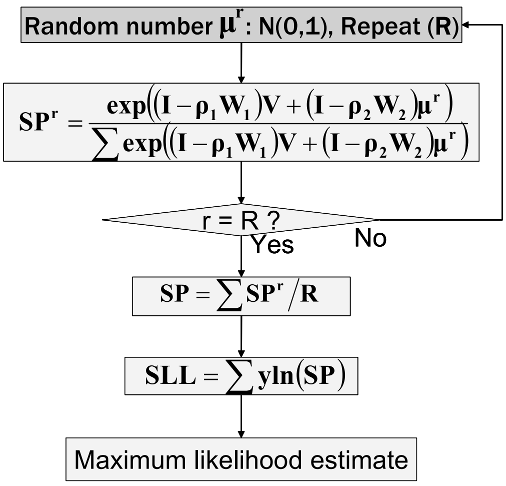

Figure 1 Maximum Likelihood Estimation for MLADE model

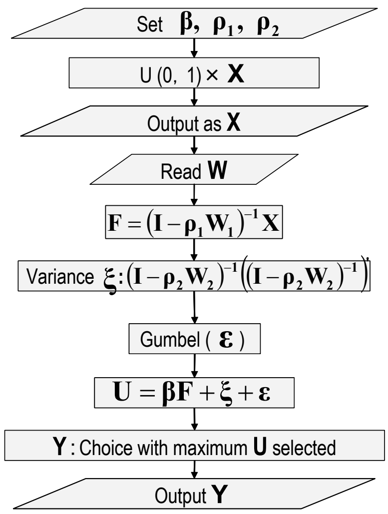

Figure 2 Data Generator for Simulation

3.4 *Model Estimation*

The choice probability of the MLADE model proposed is not closed so that analytical solution is not possible. It is, however, approximated by employing the Monte Carlo simulation technique, shown in Figure 1. The method obtained substantial attention in recent years (see, for example, Bhat *et.al.*, 2003). The choice probabilities are approximated by averaging over the R numbers of simulated probability, shown in (6). The simulated log-likelihood function is constructed as (7) where **y** is the choice results. All estimations were carried out using the GAUSS programming language.

$$\mathbf{SP} = \frac{1}{R}\sum_{r=1}^{R} \frac{\exp\left(\left(\mathbf{I}-\boldsymbol{\rho}_1\mathbf{W}_1\right)^{-1}\boldsymbol{\beta}\mathbf{X}+\left(\mathbf{I}-\boldsymbol{\rho}_2\mathbf{W}_2\right)^{-1}\mathbf{u}^r\right)}{\sum\exp\left(\left(\mathbf{I}-\boldsymbol{\rho}_1\mathbf{W}_1\right)^{-1}\boldsymbol{\beta}\mathbf{X}+\left(\mathbf{I}-\boldsymbol{\rho}_2\mathbf{W}_2\right)^{-1}\mathbf{u}^r\right)} \tag{6}$$

$$\mathbf{SLL} = \sum_{\mathbf{n}} \mathbf{y}\ \mathbf{ln(SP)} \tag{7}$$

4. Analysis with the Simulated Data

To validate the proposed model, this section presents an analysis with artificial data that was made by simulation, shown in Figure 2. Appropriate spatial effects were obtained by specifying the spatial effect parameters in (2) and (4) as β=6, ρ_1=0.2, and ρ_2=0.8. We evaluate four models discussed previously (L, LAD, MLAE, and MLADE models) by comparing the values specified to the parameter estimates obtained. Table 1 shows the estimation result of the four models. Generally, all the three models with spatial effects incorporated have better performance than the standard logit (L) model, i.e., closer parameter estimates to the pre-specified value and higher likelihood ratio. More specifically, our MLADE model, which incorporates the spatial effects for the deterministic and error components, has gaining performance over the LAD and MLAE models. It is therefore judged that despite of its complicated and expensive calculation, MLADE model is worth for real application in location choice analysis.

Table 1 Parameter Estimation from the Simulated Data

	L	LAD	MLAE	MLADE
Explanatory β	3.994 (21.69)	4.128 (21.586)	5.010 (18.68)	5.724 (18.07)
Spatial Interaction ρ 1	—	0.084 (3.28)	—	0.175 (5.90)
Spatial autocorrelation ρ 2	—	—	0.632 (9.04)	0.761 (25.87)
Adjusted likelihood ratio	0.252	0.255	0.253	0.263

5. A Case Study of Location Choice in Sendai

In the empirical analysis of this paper, we apply the four models to a case study in Sendai City of Japan. The problem is location choice of residents to the specific four zones in Sendai City, shown in Figure 3. Three of these zones are located close to each others in the south while another zone is located separately in the north. The three zones entail strong spatial effects while another entails weak spatial effects to location choice. The data used in the analysis includes travel survey questionnaire in 1996, basic land-use survey in Sendai Metropolitan Area in 1992, and land price data in 1992. The explanatory variables include number of commercial building, distance to transportation facility, income heterogeneity, family size heterogeneity, and vehicle ownership heterogeneity.

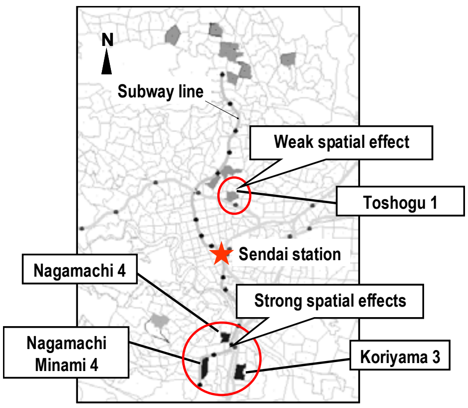

Figure 3 Location Choices of Four Zones in Sendai City of Japan

5.1 *Spatial Weight Matrices*

The models in (2) and (4) require the specification of spatial weight matrices $\mathbf{W_1}$ and $\mathbf{W_2}$, which are differently specified in the literatures, e.g., negative exponential, inverse distance, etc. This paper presents the sensitivity analysis of the functional form of $\mathbf{W_1}$ and $\mathbf{W_2}$. Figure 4 and 5 show the results for the deterministic component and error component respectively. The functional form that gives high significance of parameter estimates ρ and high likelihood ratio is selected for (2) and (4) as $w_{ij} = \exp\left(-d_{ij}^{1.5}\right)$ for $\mathbf{W_1}$ and $w_{ij} = 1/d_{ij}^2$ for $\mathbf{W_2}$, respectively.

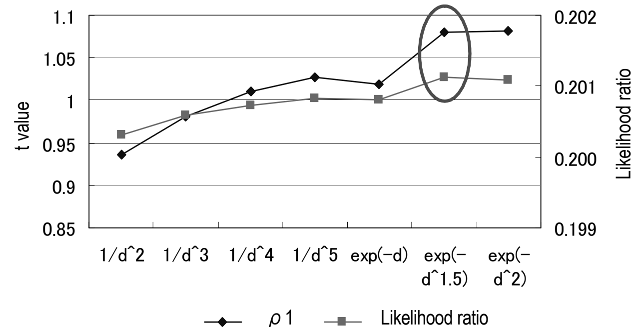

Figure 4 Spatial Weight for the Deterministic Terms (W_1)

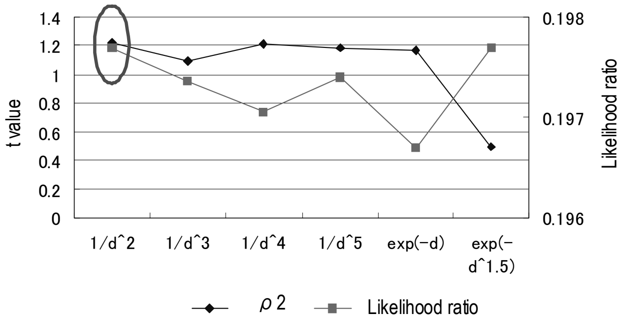

Figure 5 Spatial Weight for the Error Terms (W_2)

5.2 *Estimation Results*

Stable estimation is achieved at about 700 simulations. Table 2 shows the result. By considering the likelihood ratio the proposed model (MLADE) has the best performance as expected. Although the improvement amount is small, it is enough to say that structuralized specification of spatial effects has significant role to improve the result.

Table 2 Results of the Parameter Estimation

	L	LAD	MLAE	MLADE
Number of commercial building	0.028 (2.09)	0.037 (2.39)	0.037 (1.92)	0.052 (1.44)
Distance to the nearest transportation facility	-6.949 (-4.28)	-6.516 (-3.94)	-7.619 (-3.96)	-6.904 (-3.20)
Land price (1,000 yen/m2) / Income (100,000 yen/year)	-2.036 (-2.84)	-1.889 (-2.59)	-2.279 (-2.77)	-2.104 (-2.47)
Difference of household income and the zone-average income	-0.826 (-4.81)	-0.849 (-4.82)	-0.972 (-4.48)	-1.046 (-2.91)
Difference of the number of member in a family and the zone-average number	-1.513 (-1.85)	-1.466 (-1.76)	-1.868 (-1.89)	-1.890 (-1.72)
Difference of the number holding vehicle and the zone-average number	-4.903 (-5.36)	-4.622 (-4.99)	-5.683 (-5.21)	-5.343 (-5.05)
Spatial Interaction	—	0.013 (1.08)	—	0.015 (1.06)
Spatial Autocorrelation	—	—	0.442 (0.71)	0.608 (1.10)
Adjusted likelihood ratio	0.188	0.191	0.190	0.193

6. Concluding Remarks

This paper has shown that spatial effects can be substantially accommodated by the structuralized specification of the utility function. The application to a case study has validated the proposed model. Moreover, the better model performance could be expected by using an improved set of data and left for further study.

Acknowledgements

The authors would like to express their sincere appreciation to Dr. Daisuke FUKUDA of Tokyo Institute of Technology for kindly providing the basic information on the topic.

References

Anselin, L. (1988) Spatial Econometrics, Methods and Models. Boston: Kluwer Academic.

Ben-Akiva, M. and Bolduc, D. (1996) Multinomial Probit with a Logit Kernel and a General Parametric Specification of the Covariance Structure. Department d'Economique, Universite Laval, Quebec, Canada.

Bhat, C. and Guo, J. (2003) A Mixed Spatially Correlated Logit Model: Formulation and Application to Residential Choice Modeling. Transportation Research B, Vol.37, forthcoming.

McFadden, D. and Train, K. (2000) Mixed MNL Models for Discrete Response. J. of Applied Econometrics, 15, 5, 447-470.

McMillen, D.P. (1992) Probit with Spatial Autocorrelation. Journal of Regional Science 32, 335-348.