EFFECTS OF SPATIAL CORRELATION ON THE CHOICE BEHAVIOR OF URBAN GOODS DISTRIBUTION*

By Wisinee WISETJINDAWAT**, Kazushi SANO***, and Shoji MATSUMOTO****

1. Introduction

Distribution model is an important part of the traditional four-step approach for modeling freight transportation. Distribution models can be categorized into two main types: trip-based and commodity-based approaches. Trip-based approach, just like in passenger demand model, distributes the total generated trip attraction and generation to each OD pair. On the contrary, commodity-based approach focuses on the principle of freight movement which says that the freight movement is the result of the movement of commodities. In stead of generating trips, commodity-based approach estimates the level of commodity production and consumption and distributes the generated commodities to each OD pair at the commodity distribution stage. Generally, gravity model is utilized for modeling the distribution patterns of freight movement for both trip-based and commodity-based approaches. However, the typical gravity model still lacks of considering the fundamental behavior in which commodity distribution patterns are originally decided by the demand for consumption of the commodities at the destination. A fractional split distribution model developed by Sivakumar and Bhat[1)] follows this concept and has been applied to statewide freight movement in the United States. The model was a discrete choice paradigm to estimate the fraction of freight amount produced by each zone origin to be consumed at each destination zone.

As computer technology has been dramatically developed and, as of recently, has reached a very high performance level, it is becoming more practical to develop freight models at the microscopic level. Microscopic models deal with each individual behavior instead of dealing with a zone as in macroscopic models. The individual can be a truck, an office, or a company depending on the level of modeling considered in each model. The current freight demand modeling tends to the microscopic level as it results in a more realistic and policy sensitive model comparing with the macroscopic one.

Flexibility in specifications of mixed logit model is an attractive characteristic and found suitable for modeling the spatial interactions in purchasing choice behavior. Mixed logit model is currently gaining popularity in behavioral science. In recent years, a number of mixed logit models have been developed in several topics, for example, transportation mode choice, residential choice and so on. Spatial mixed logit on residential choice is presently an interest of many researchers and can be found in the works of Bhat and Guo[2)], Miyamoto et al[3)], and Mohammandian et al[4)]. This study utilizes mixed logit framework with the specifications of the disturbance terms to consider the spatial interactions in modeling of commodity distribution.

A commodity distribution model proposed here is a commodity-based model incorporating supply chain characteristics and spatial effects, which considers supply chain structure and the spatial correlations across each individual choice decision on purchasing commodities from each supplier. The proposed model is a disaggregate type which covers the first two steps of the traditional four-step approach in which commodity generation and distribution are undertaken.

2. Model Structure

(1) Commodity Distribution Model

Firstly, the model generates the total amount of commodity produced and consumed by each firm. The amounts of production and consumption are estimated from the firm size indicators such as number of employees and floor area.. Commodity production and consumption are estimated by type of commodity as shown in (1) and (2).

$$G_i^k = f(x_1, x_2, ..., x_n) \tag{1}$$

$$A_j^k = f(x_1, x_2, ..., x_n) \tag{2}$$

Where, G_i^k is the amount of commodity k produced by firm *i*. A_j^k is the amount of commodity *k* consumed by firm *j*.

The generated commodities are linked from consumption firm to production firm according to the attractiveness of suppliers and relationship between firms in supply chain. The attractiveness of suppliers is derived from supplier location and the amount of commodity produced by that supplier. Distribution channels are the paths connecting the customers and shippers of a commodity in a supply chain. Therefore, the distribution model consists of three parts:

* Keywords: Physical distribution, Spatial correlation, Mixed logit, Urban goods movement
** Graduate Student, M. Eng., Dept. of Civil and Environmental Eng., Nagaoka University of Technology,
(1603-1 Kamitomioka-machi, Nagaoka, Niigata, 940-2188 Japan, E-mail: wisinee@stn.nagaokaut.ac.jp)
*** Member of JSCE, Associated Professor, D. Eng., Dept. of Civil and Environmental Eng., Nagaoka University of Technology,
(1603-1 Kamitomioka-machi, Nagaoka, Niigata, 940-2188 Japan, E-mail: sano@vos.nagaokaut.ac.jp)
**** Member of JSCE, Professor, D. Eng., Dept. of Civil and Environmental Eng., Nagaoka University of Technology,
(1603-1 Kamitomioka-machi, Nagaoka, Niigata, 940-2188 Japan, E-mail: shoji@vos.nagaokaut.ac.jp)

probability of distribution channel, probability of shipper location, and probability of selecting each shipper. The product of all three parts yields the probability of a given shipper being selected. The mathematical form of the model can be shown as follows:

$$P^k(z) = P(C^k) \cdot P(z|C^k) \cdot P(i|C^k, z) \quad i \in C^k, z \tag{3}$$

where, $P^k(i)$ is the probability of shipper i being selected for commodity k. $P(C^k)$ is the probability of distribution channel C being used for commodity k. $P(z|C^k)$ is the probability of zone z being selected, given distribution channel C. $P(i|C^k, z)$ is the probability of shipper i is selected, given distribution channel C and zone z

The first part, the distribution channel probability, $P(C^k)$ is calculated directly from the empirical data as the percentage of firms selecting to purchase a commodity from each industry type of shipper. The second part, the location probability, $P(z|C^k)$ is viewed as a choice selection problem with multiple alternatives. A customer is assumed to select the zone that maximizes their utility function among the zone alternatives. This study utilizes a spatial mixed logit model which incorporates the spatial interaction among zone alternatives. The conditional probability that zone z is chosen, given distribution channel C^k, is expressed below:

$$P(z|C^k) = \int_\beta \left(\frac{e^{X_{zn}\beta + F_{zn}T\zeta_n}}{\sum_{z' \in Z} e^{X_{z'n}\beta + F_{z'n}T\zeta_n}} \right) n(\zeta, I_M) d\zeta \tag{4}$$

The third part, the shipper probability given distribution channel C^k and zone z is to identify the shippers from which a customer purchases. However, due to the limitation of the survey data that does not identify the exact purchased shipper; we then assume that the shipper probability can be derived directly from the proportion of the size indicators of each shipper, which in this case is the amount of commodity production. The conditional probability that shipper i is chosen, given C^k and z, is as shown below:

$$P(i|C^k, z) = \frac{\exp(G_i^k)}{\sum_{i' \in i} \exp(G_{i'}^k)} \tag{5}$$

where, G_i^k is the production amount of commodity k from shipper i

Once the shipper probability ($P^k(i)$) is determined, random numbers are generated and used with the shipper probability to indicate the selected shippers and determine the share for each selected shipper. Commodity flows are derived from the product of shipper share and consumption amount for each firm as shown in (6).

$$Q_{ij}^k = P^k(i) \cdot A_j^k \tag{6}$$

where, Q_{ij}^k is monthly commodity flow between firm i to firm j for commodity . $P^k(i)$ is probability of shipper i is selected for commodity k. A_j^k is monthly consumption amount consumed by firm j for commodity k.

(2) Spatial interactions in zone choice probability

Spatial heterogeneity among people preferences is a key role in decision choice behavior since the preferences among individuals are correlated. Consumer's decision is impacted by spatial interactions with other consumers who live nearby. Spatial Interactions are incorporated into the proposed model in the part of zone choice probability. To incorporate the spatial interaction, mixed logit model is used to explain choice decision process since its flexible structure permits several specifications of the disturbance. The flexibility of mixed logit model is caused by the disturbance term which contains a flexible probit-like term and an additive iid extreme value term. According to Ben-Akiva et al[5], vector of utility of alternatives for decision maker n is taken in the general form of mixed logit model as follows:

$$\mathbf{U}_n = \mathbf{X}_n\boldsymbol{\beta} + \mathbf{F}_n\mathbf{T}\boldsymbol{\zeta}_n + \mathbf{v}_n \tag{7}$$

where, $\boldsymbol{\beta}$ is a [K×1] vector of unknown parameters. $\mathbf{X}_n$ is a [J×K] matrix of explanatory variables for decision maker n. $\mathbf{F}_n\mathbf{T}\boldsymbol{\zeta}_n$ is the covariance structure among alternatives zones. $\mathbf{F}_n$ is a [J×M] matrix of factor loading. $\mathbf{T}$ is a [M×M] matrix of standard deviation where σ on the diagonal. $\boldsymbol{\zeta}_n$ is a [J×1] vector of random numbers follows $N(0,1)$. $\mathbf{v}_n$ is a [J×1] vector of iid gumbel random variables.

The probability of alternative z being selected for decision maker n is the integrals of standard logit probability over a density of parameters. P_{nz} is formulated as shown in (8) in which $n(\zeta, I_M)$ is the joint density function of ζ.

$$P_{nz} = \int_\zeta \Lambda(z|\zeta) n(\zeta, I_M) d\zeta = \int_\beta \left(\frac{e^{X_{zn}\beta + F_{zn}T\zeta_n}}{\sum_{z' \in Z} e^{X_{z'n}\beta + F_{z'n}T\zeta_n}} \right) n(\zeta, I_M) d\zeta \tag{8}$$

Since the integral form cannot be calculated, the probability is approximated through the simulation of any given value of ζ_n^d, where r denotes draw of n from the distribution of ζ. The average of these probabilities yields the simulated probability as shown as follows:

$$\hat{P}_{zn} = \frac{1}{R} \sum_{r=1}^{R} \Lambda(i|\zeta_n^r) \tag{9}$$

Likewise, the simulated probability inserted into the log-likelihood function results in the simulated log-likelihood function as follows:

$$SLL = \sum_{n=1}^{N} \sum_{z \in Z} y_{zn} \ln \hat{P}_{zn} \tag{10}$$

where, y_{zn} is the fraction of commodities consumed by customer n supplied from zone z. For each customer, the total consumption fractions from all zones must always be one and must satisfy the following constraints.

$$0 \le y_{zn} \le 1 \text{ and } \sum_{z \in Z} y_{zn} = 1 \tag{11}$$

(3) Incorporating the correlation among alternatives in the error term

The correlations among alternatives can be taken into account by specification of the disturbance term. McMillen[6] inserted the spatial autoregressive into the disturbance term of probit model. Ben-Akiva et al[5] suggested the generalized autoregressive to enter into the disturbance term of mixed logit. This research therefore utilizes the generalized autoregressive to explain the spatial correlation among alternatives as defined as follows:

$$\boldsymbol{\xi}_n = \rho \mathbf{W} \boldsymbol{\xi}_n + \mathbf{T}_n \boldsymbol{\zeta}_n = (\mathbf{I} - \rho \mathbf{W})^{-1} \mathbf{T}_n \boldsymbol{\zeta}_n \tag{12}$$

where, ρ is a scalar unknown parameter. $\mathbf{I}$ is a [J×J] identity matrix. $\mathbf{W}$ is a [J×J] weight matrix identifying the correlation among alternatives (zones). $\mathbf{T}_n$ is a [J×J] matrix which has standard error (σ) on the diagonal. $\boldsymbol{\zeta}_n$ is a [J×1] matrix of random variables which follow normal distribution $N(0,1)$.

The spatial correlation either strong or low depends on spatial locations among alternative zones. Typically, the distance between zones is an indicator to measure the level of correlation. The weight matrix ($\mathbf{W}$) represents the level of correlations among alternative zones and assumed to be a function of distance between zones. The correlation between zones i and j is:

$$w_{ij} = \frac{1}{d_{ij}^{\alpha}} \tag{13}$$

where, w_{ij} is an element of weight matrix $\mathbf{W}$. d_{ij} is the distance between zones i and j. α is an unknown parameter.

Substituting $\boldsymbol{\xi}_n$ into (7), the simulated probability becomes:

$$\hat{P}_n = \frac{1}{R} \sum_{r=1}^{R} \sum_{z \in Z} y_z \frac{\exp[\mathbf{X}_n \boldsymbol{\beta} + (\mathbf{I} - \rho \mathbf{W})^{-1} \mathbf{T} \boldsymbol{\zeta}_n^r]}{\sum_{z \in Z} \exp[\mathbf{X}_n \boldsymbol{\beta} + (\mathbf{I} - \rho \mathbf{W})^{-1} \mathbf{T} \boldsymbol{\zeta}_n^r]} \tag{14}$$

3. Empirical Results

The case study is carried out on the Tokyo Metropolitan Area (TMA) which covers the area of five prefectures including Tokyo, Kanagawa, Chiba, Saitama, and the southern part of the Ibaraki. The study area is divided into 56 zones according to the A-zone classification of the Tokyo Metropolitan Goods Movement Survey (TMGMS). The area comprises 52 zones within the study area and 4 zones for the prefectures near the study area for analysis of the external trips. The TMGMS data classifies commodities into 50 types which we categorize into 8 groups including 1) Agricultural Products, 2) Forestry Products, 3) Mineral Products, 4) Metal and Machinery Products, 5) Chemical Products, 6) Light Industry Products, 7) Other Products, and 8) Wastes and Scraps. In the same way, the industry types are grouped into 13 groups including 1) Agriculture, Forestry, and Fishery, 2) Mining, 3) Construction, 4) Chemical Manufacturer, 5) Metal Manufacturer, 6) Machinery Manufacturer, 7) Other Manufacturer, 8) Material Wholesaler, 9) Product Wholesaler, 10) Retailer, 11) Warehouse, 12) Electricity, Gas, and Water Supplier, and 13) Service and Government Work.

Firstly, the probability of distribution channel being selected, $P(C^k)$ is determined directly from the survey data by the percentage of commodity weight being purchased from each industry type of shipper by each industry type of customer for each of the commodity types. Secondly, the model to determine the probability of zone being selected $P(z|C^k)$ is developed separately for each commodity type that is purchased by each industry type of customer. The independent variables considered in this analysis consist of two major variable types: zonal impedance variables and zonal attraction variables. The distance between zones is used as the zonal impedance variable in this study. The zonal attractiveness variables include the amount of commodity production, number of establishments, population, and number of employments. Spatial weight matrix, $\mathbf{W}$ represents the spatial relationship among the alternative zones. Distance between zones is usually used to measure contiguity relationship between alternative pair. The weight matrix in this study is defined using a distance-decay function as shown in (13). The distance used here is the average travel distance between zones which is calculated from the empirical data.

The model for probability of a zone being selected is separately developed for each commodity type that is purchased by each industry type of customer from each industry type of shipper. An example of the model is selected to be discussed in this paper which is zone choice model for manufacturers (*Ic7*) purchasing light industry products (*k6*) from other manufacturers (*Is7*). Model estimation is performed using GAUSS maximum likelihood estimator. The comparing results between with and without spatial correlation are presented in Table 1. The t-statistics of the zonal attractiveness variables are statistically significant for both models. The sign of the variables indicates the customer's preference for the proportion of commodity being purchased. As expected, the positive sign of the zone attractiveness variables implies that a firm prefers purchasing a larger quantity from the zones having the larger number of generated commodities and establishments. Likewise, the zones, which have the higher attractiveness in terms of number of population and number of employees, receive the larger market share. The spatial parameters are statistically significant in term of t-value and improve the value of adjusted log-likelihood ratio confirming that spatial interaction among zone alternatives affects to the decision process of a customer on purchasing a commodity.

Table 1. Estimated parameters for light industry products purchased by manufacturers from other manufacturers

Variables	**MNL**		**SML**	
	Parameter	**t-value**	**Parameter**	**t-value**
Zonal attractiveness variables				
Amount of Generated Commodities (in 1,000,000 kg.)	0.0108	26.90	0.0153	15.36
Number of Establishments (in 1,000's.)	0.0383	4.52	0.0578	5.29
Population (in 10,000 persons.)	0.0030	2.48	0.0026	2.58
Density of Number of Employees (in 10,000 persons/km^2)	0.0588	1.39	0.1609	3.08
Impedance variables				
Distance (km)	-0.0249	-16.72	-0.0337	-14.50
Correlation Variables				
γ			2 [a)]	
ρ			-11.673	-1.63
σ			2.8734	4.04
Number of observations	667		667	
Log-likelihood at zero	-2635.5		-2635.5	
Log-likelihood at convergence	-2118.5		-2024.9	
Adjusted Log-likelihood ratio	0.196		0.232	

Note: [a)] Parameter is constrained to 2 for identification purpose

4. Conclusion

This paper has proposed an alternative approach for modeling the commodity flows in urban area. The proposed model attempts to explain the fundamental of commodity movement in which the movement of freight is an outcome of commodities flow passing trough supply chains. Each individual consumption point is linked to each production point according to the attractiveness of each production point resulting in commodity flows between consumption and production points. Fundamentally, it is the demand at consumption points that determines the amount of commodity that should be supplied at production points. Commodities are distributed from firm to firm over an area according to their relationship in a supply chain.

The proposed model was applied to urban freights movement in the Tokyo Metropolitan Area. The comparing results with and without spatial correlation demonstrates that the spatial interaction has significant effects on the decision process. The proposed commodity distribution model incorporates complex relationship among freight decision makers interacting in supply chains in which the spatial interaction is considered.

References

1) Sivakumar, A. and Bhat, C.: Fractional Split-Distribution Model for Statewide Commodity-Flow Analysis. Transportation Research Record 1790, pp. 80-88, 2002.
2) Baht, C.R. and Guo, J.: A Mixed Spatially Correlated Logit Model: Formulation and Application to Residential Choice Modeling, Transportation Research Part B, Vol. 38, pp. 147-168, 2004.
3) Miyamoto, K. *et al*: Discrete Choice Model with Structuralized Spatial Effects for Location Analysis, Proceeding of the 83rd Annual Transportation Research Board, 2004.
4) Mohammadian, A. *et al*: Incorporating Spatial Dependencies in Random Parameter Discrete Choice Models, Proceeding of the 84th Annual Transportation Research Board, 2005.
5) Ben-Akiva, M. *et al*: Specification, Identifications, and Estimation of the Logit Kernel (or Continuous Mixed Logit) Model, 2001.
6) McMillen, P.: Probit with Spatial Autocorrelation, Journal of Regional Science, Vol. 32, No. 3, pp. 335-348, 1992.