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

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

1. Introduction

A distribution model in the traditional four-step approach is an important part for demand modeling of freight transportation. Modeling in freight transportation generally can be categorized into two main types: trip-based and commodity-based approaches[1]. Trip-based approach, like in passenger demand model, distributes the total generated trip attraction and generation to each OD pair. On the contrary, commodity-based approach pays an attention to the fundamental of freight movement 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. Generally, gravity model is utilized for modeling the distribution patterns of freight movement for both trip-based and commodity-based approaches. Gravity model calculates the flows between each pair of origin and destination zones weighted by an impedance term which the distance between zones is mostly used. 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[2] follows this concept and has been applied to statewide freight movement in the United States. The model used discrete choice paradigm to estimate the fraction of freight amount produced by each zone origin to be consumed at each destination zone. The results confirmed that discrete choice model can produced better performance than gravity model.

(1) Spatial Interactions among Freight Agents

Freight movement is, in fact, the results of the complex interaction among freight agents (including shippers, customers, and carriers). Model that considers the interaction is necessary for modeling freight movement to be more realistic. For commodity distribution problem, it can be viewed as the interplay between a shipper and a customer of a commodity and finally results in the commodity movements. In marketing science, firms that gather together will gain benefit such as lower cost or the greater market share. This characteristic should be considered into modeling of freight movement. However, models that consider spatial interaction mostly focus on housing choice problem or in marketing science. The examples of spatial interaction choice model in housing choice problem can be found in the works of McMillen[3], Bhat and Guo[4], Miyamoto et al[5], and Mohammadian[6]. Various spatial models in marketing science are reviewed in the paper of Bradlow et al[7]. Spatial interaction plays an important role in customer's decision making since customer while making a decision on purchasing a commodity does not act in isolation but, in fact, is influenced by the others who located nearby. There are only few researches incorporate the spatial interaction into freight transportation modeling. The spatial commodity distribution model proposed by Wisetjindawat et al[8] is an example of this modeling type. The model assumed that the commodity flows are originated by the demand of consumption at the consumption point that should be supplied at the generation point. The model is a microscopic modeling type that considers the individual behavior of a customer on purchasing a commodity from each of the suppliers. The model compared the effects of the spatial interactions between the interaction among alternatives and the interaction among decision maker themselves by using mixed logit framework specified the deterministic and error parts to consider the interactions among decision makers and among alternatives respectively. The results indicated that decision making are mainly influenced by the interaction of the decision maker with the others. However, due to complexity of the

**Graduate Student, M. Eng., Department. of Civil and Environmental. Engineering, Nagaoka University of Technology
   (1603-1 Kamitomioka-machi, Nagaoka, Niigata, 940-2188 Japan, E-mail: wisinee@stn.nagaokaut.ac.jp)
***Member of JSCE, Ass. Prof., D. Eng., Department. of Civil and Environmental. Engineering, Nagaoka University of Technology,
   (1603-1 Kamitomioka-machi, Nagaoka, Niigata, 940-2188 Japan, E-mail: sano@ nagaokaut.ac.jp)
****Member of JSCE, Prof., D. Eng., Department. of Civil and Environmental. Engineering, Nagaoka University of Technology,
   (1603-1 Kamitomioka-machi, Nagaoka, Niigata, 940-2188 Japan, E-mail: shoji@nagaokaut.ac.jp)

modeling which requires long calculation time, mixed logit model is incapable to be applied to analyze the commodity distribution in the entire area of the Tokyo Metropolitan Area. For example, the calculation time takes more than 5 hours for 500 samples. This paper therefore attempts to simplify to model and focuses mainly on the differences among commodity types and industry types, while maintains the important characteristics in which the spatial interaction among decision makers is incorporated.

(2) Objective of the Paper
This paper is an extension of the above model which is a commodity-based model incorporating supply chain characteristics and spatial effects. This paper utilizes a logit model framework to consider the spatial correlations across each individual choice decision on purchasing commodities from each supplier. This paper utilizes spatial multinomial logit model since the results of the previous paper indicated that the interaction among customers significantly improve the model performance comparing with the interaction among alternatives and the calculation of multinomial logit requires less calculation time and possible to apply to the goods movement overall the entire study area. This paper is structured into four parts. Next section will provide model framework, mathematical formulations. The results of the model applied to urban goods movement in the Tokyo Metropolitan Area will be presented afterwards. Finally, the discussion and recommendation on the proposed model will be provided in the conclusion part.

2. Model Structure

The proposed model is a disaggregate model type, which considers the behavior of each individual firm. The model covers the first two steps of the traditional four-step approach in which commodity generation and distribution are undertaken.

(1) Commodity Distribution Model
First, 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).

$$G_i^k = f(x_{i1}, x_{i2}, ..., x_{in}) \tag{1}$$

$$A_j^k = f(y_{j1}, y_{j2}, ..., y_{jm}) \tag{2}$$

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.

Then, the generated commodities are distributed from firm to firm over the study area. Originating from the consumption firm, the firm selects supplier's firms for each commodity according to the attractiveness of the suppliers and relationship between the firms in the supply chain of commodity. The attractiveness of suppliers is derived from supplier location apart from customer location and the amount of commodity produced by the supplier. Distribution channels express the relationship between the customers and shippers of a commodity in a supply chain. Therefore, the distribution model consists of three parts: 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:

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

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

First, the distribution channel probability, $P_j(C^k)$ is calculated directly from the empirical data. From the data, we can estimate the percentage of firms selecting to purchase a commodity from each industry type of shipper which is used to represent distribution channel probability. Second, the location probability, $P_j(z|C^k)$ 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 multinomial logit model which incorporates the spatial interaction among each individual customer. The details on this model will be provided in the next section.

Third, 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:

$$P_j(i|C^k, z) = \frac{\exp(G_i^k)}{\sum_{i' \in \square} \exp(G_{i'}^k)} \tag{4}$$

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

Once the shipper probability ($P_j^k(i)$) 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 (5).

$$Q_{ij}^k = P_j^k(i) \cdot A_j^k \tag{5}$$

where, Q_{ij}^k is monthly commodity flow between firm i to firm j for commodity k. $P_j^k(i)$ 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) Zone Choice Probability with Spatial Interactions, $P_j(z|C^k)$

The preferences among individuals are correlated; a consumer's decision is often influenced by spatial interactions with other consumers who live nearby. Spatial interactions are incorporated into the proposed model in the zone choice probability. The model predicts the probability of consumption by each customer to purchase a commodity originating from each production zone. Each individual firm is assumed to choose a zone for purchasing a commodity according to the attractiveness of each zone, represented by the utility function of alternative zones. The sum of the fraction across all production zones for each customer's firm must equal one, and the fraction should satisfy the following constraints.

$$\sum_{z \in Z} y_{zj} = 1 \quad \text{and} \quad 0 \leq y_{zj} \leq 1 \tag{6}$$

Where, y_{zj} is the fraction of commodity consumed by customer j and supplied from zone z. Z is the total number of zone alternatives. Therefore, the probability that zone z, given distribution channel C^k, is selected for purchasing commodity k for customer's firm j can be expressed as follows:

$$P_j(z|C^k) = \frac{\exp(V_{zj})}{\sum_{z \in Z} \exp(V_{z'j})} \tag{7}$$

Where, V_{zj} is the zonal attractiveness of zone z to customer j.

The model incorporated spatial dependences proposed by Mohammadian et al[6] to explain housing choice behavior considers the interactions among decision makers by adding the interaction part into the deterministic part of the choice model. The application of the model shows an interesting improvement of the results with the spatial interaction over those without the spatial interaction. The spatial commodity distribution model developed by Wisetjindawat et al[8] confirms that customer decision making is significantly influenced by the decision of the other customers. We, therefore, apply this concept to explain the interaction among customers. The specification of distance decay functions utilized in this paper is an inverse distance function. For the deterministic part (V_{zj}), the model after adding the interactions becomes

$$V_{zj} = X_{zj}\beta + \phi_{zj} = \sum_{k=1}^{K} \beta_{zk} x_{zjk} + \lambda \sum_{s=1}^{S} y_{zs} \frac{1}{d_{js}^\delta}, \tag{8}$$

where β_{zk} is a parameter corresponding to the observed characteristic x_{zjk} of alternative zone z and customer j. λ is a scalar unknown parameter. y_{zs} is the consumption fraction of alternative zone z of customer s when S is the total number of customers. d_{js} is the distance between customers j and s, where customer j is the decision maker and customer s is another customer. δ is a scalar unknown parameter.

The estimation of the model is as same as the method for general multinomial logit by using the maximum likelihood method. The paper performs model calibration using the GAUSS programming language.

3. Empirical Results

(1) Data Set
The Tokyo Metropolitan Area (TMA) is selected for analysis in this paper. The area covers five prefectures including Tokyo, Kanagawa, Chiba, Saitama, and the southern part of Ibaraki. The area is divided into 56 zones according to the A-zone classification of the Tokyo Metropolitan Goods Movement Survey (TMGMS) comprising 52 zones within the study area and 4 zones for the prefectures near the study area for analysis of the external trips. The model calibration uses the data of TMGMS survey collected from the firms throughout the TMA by the City and the Regional Development Bureau of the Japanese Ministry of Construction in 1982. The data consist of records on commodity movement and truck movement of each firm. The data were from approximately 46,000 firms, corresponding to three percent of all firms in the study area. We use the database of 1982 since this database provide the most complete information required for model calibration. We categorize the commodities 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. Agricultural products include all unprocessed foods. Forestry products are wood, timber, charcoals, and other related products. Mineral products include natural gas, petroleum, sand, rock, and other related products. Metal and machinery products are metal materials (such as aluminum and steel) and machines. Chemical products include cement, glass, gasoline, plastic, asphalt, and other related products. Light industry products are all processes foods, paper, thread, and other related products. Other products include newspapers, magazines, cloths, furniture and daily use commodities. Wastes and scraps include all type of wastes such as industrial and residential wastes.

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.

(2) Results and Discussions
We analyses the model for all industry types and commodity types that are mentioned above. This paper compares two types of the commodity distribution model between the model using general multinomial Logit model (MNL) and the multinomial logit model incorporating spatial interactions (SMNL). The independent variables considered in this analysis consist of two major variable types: zonal attractiveness variables and zonal impedance variables. The zonal attractiveness variables include the amount of commodity production (GEN), number of establishments (NC), population (POP), zone area (AREA), and number of employments (EMP). The distance between zones (DIST) is used as the zonal impedance variable in this study which is the average travel distance calculated from the empirical data.

The MNL and SMNL models are calibrated separately for each industry type of shipper, industry type customer, and commodity type. The models are tested using statistical tests such as t-value of each parameter and ρ^2 of the model. The parameter that has t-value less than 1.96 will be omitted for each model. ρ^2 of all models, which indicate the model performance, range from 0.169 to 0.782. The results of the average value of zonal attractiveness variables classified by commodity type of SMNL model are summarized in Figure 1. GEN and AREA generally affect the amount of commodities distributed from zones for all commodity types. NC is insignificant on the decision choice for purchasing agricultural and mineral products, and for receiving waste and scraps since the factors that influence on generating these commodity types are commonly the number of population and area. All zonal attractiveness variables are significant for metal and machinery products, chemical products, light industry products, and other products. On the other hand, wastes and scraps depend only on the amount of generation and area of the zone alternative. The values of AREA of agricultural and mining products are the highest comparing with the other commodities since both

commodity types are natural products and the larger area can produce the larger amount of the products. The same reason to that the values of POP for agricultural and forestry products are highest which is because of its characteristic of natural products. The value of EMP for mineral product is highest meaning that mineral products are generated the highest amount at the same number of employments comparing with other products since the commodity weight is heavy.
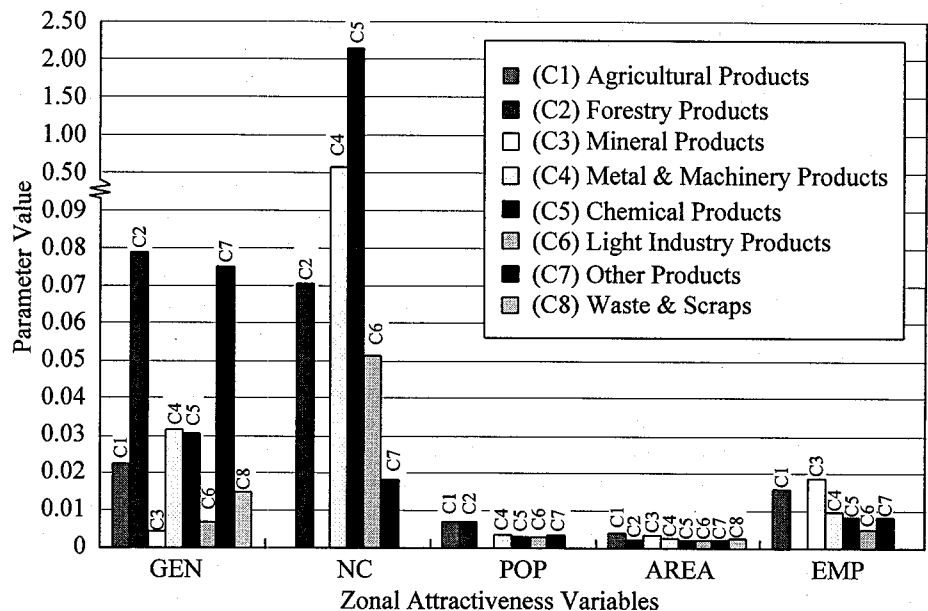

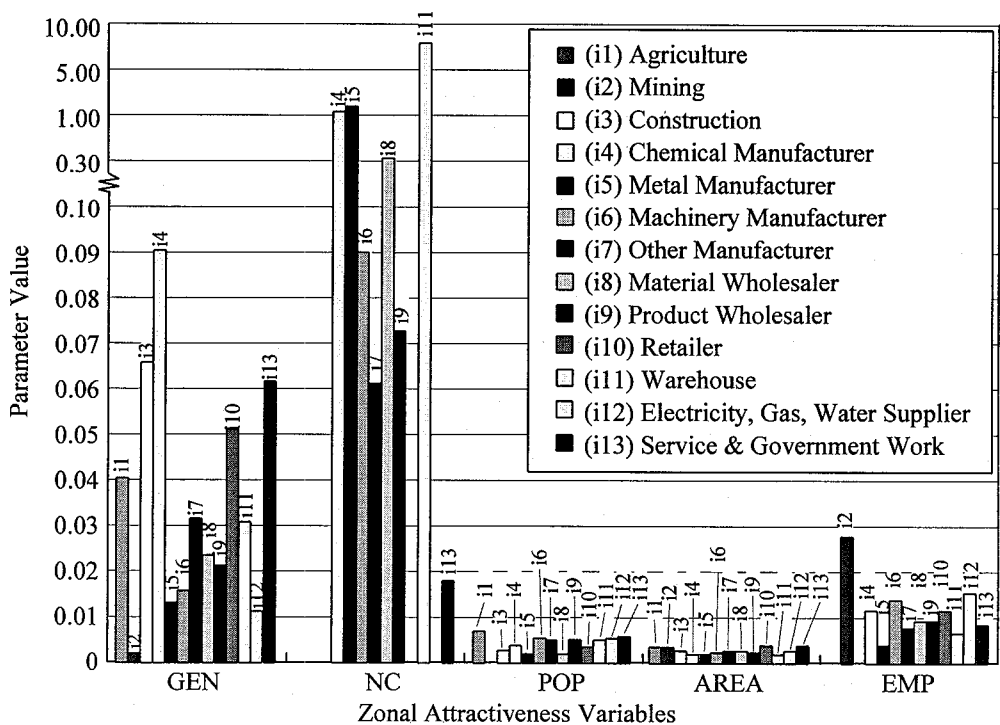Figure 1. Zonal Attractiveness Parameters Comparing among All Commodity Types

Figure 2. Zonal Attractiveness Parameters Comparing among All Industry Types

Figure 2 compares the zonal attractiveness parameters among each industry type of origin. NC and EMP in a zone have no influence on the choice decision of the customer for purchasing the commodities from agriculture industry. GEN, POP, and AREA are only the variables that impact commodity distribution from construction industry. All zonal attractiveness variables have an impact on the choice decision of customers for the commodity generated in all manufacturers, wholesalers, warehouses, and service and government work. For retailers, the commodities are generated by POP, AREA, and EMP. The value of POP for agriculture industry is the highest confirming that agricultural products are mainly generated from the number of population. AREA for retailer has the highest value since generally the density of retailers in a zone is very high and therefore causes the large amount of commodity generation from retailer. The value of EMP for mining industry is the highest; this is because of the heaviness of the commodities that the industry generate.

Figure 3 compares the average value of the distance parameter for all commodity types. The parameter for waste and scraps has the lowest value which means that waste and scraps are preferred to be transported in the shortest distance comparing with all other products because of to lower the cost of transportation. This is also similar for the case of forestry product and mineral products which are preferred to be distributed in shorter distance since the weight of products is heavy. The value for light industry products generally should be higher. This is because the value is calculated from all industry types and the share of the commodity distribution among retailers is very high in which the distribution of the products among retailer is commonly in the short distance. Figure 4 show the comparison results of the average value of the distance parameter for all industry types of the origin of the commodities. The parameter of service and government works has the lowest value and can be interpreted that the commodity originated from this industry type will be distributed in the shortest distance since this industry type generally acts as the user of the commodities. The commodities distributed from construction industry are generally conveyed in the short distance because of heaviness of the commodities. The value of distance parameter for retailer is lower than those of wholesalers and manufacturers, respectively, since it is true that the distributions of commodities from retailers are in the shorter distance. For all manufacturers and wholesalers, the value is high since the commodities are generally distributed in long distance. This is because the density of number of retailers is very high while wholesalers and manufacturers are commonly located farther away.
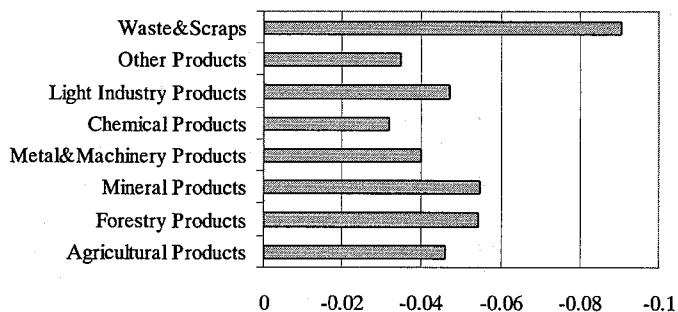

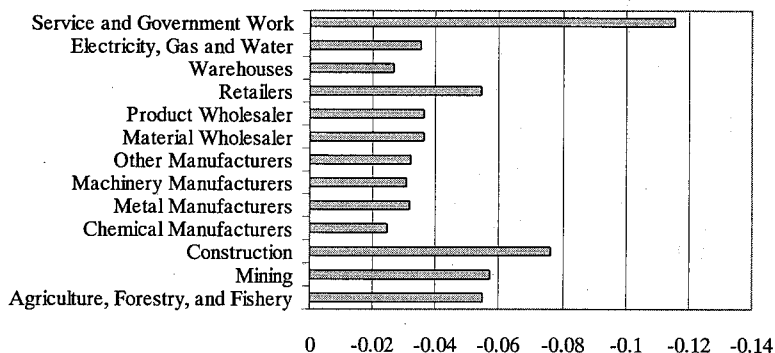Figure 3. Distance Parameters Comparing among All Commodity Types

Figure 4. Distance Parameters Comparing among All Industry Type of Origins

Figure 5 shows the validation results of fraction of commodity OD comparing between the MNL and SMNL models. The results indicate that SMNL model can significantly improve the model fitness which confirms our assumption that customers spatially interact with other customers during making purchasing decision. However, there are overestimations of the share in some zones that have the high value of zonal attractiveness variables. The overestimated zones mostly located in the central city which have the unique characteristics. This problem could be solved by setting the constraint to limit the amount of commodity generation in the zones to protect overestimation of the share in those special zones or treating the zones differently from the others. This topic is remaining for consideration in the future improvement.
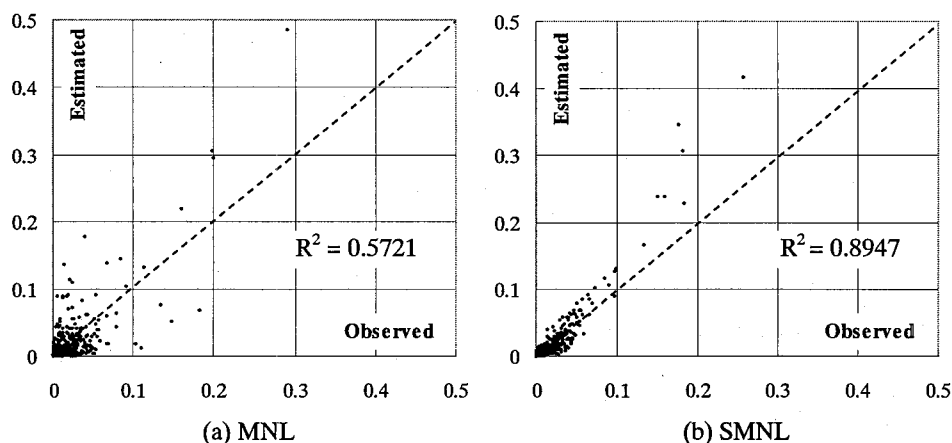

(a) MNL   (b) SMNL

Figure 5. Observed and Estimated Fraction of Commodity Flows for All Commodity Types Comparing between MNL and SMNL Model

4. Conclusion

The proposed commodity distribution model attempts to explain the fundamental of freight movement in which the movement of freight is an outcome of commodities flow passing through several freight agents in supply chains. The commodity flows are originated from the customer of the commodity; the demand of consumption is linked to each production point according to the attractiveness of each supplier. The interaction between customer and supplier of commodity results in the flow of the commodity between firms according to their relationship in a supply chain.

The proposed model incorporates not only the complex relationship among freight agents interacting in supply chains, but also the spatial interaction among freight agent. The decision of the customer of commodity is assumed to be affected by the decision of the others. The proposed model was successfully applied to urban freights movement in the Tokyo Metropolitan Area. The comparing results with and without spatial interaction demonstrates that the spatial interaction has significant effects on the decision process.

The present model, however, still not yet considers some aspects such as the choice of distribution channel since firms can change the utilization distribution channel. In addition, we could not calibrate the choice model for shipper choice because of the limitation of data that does not provide the information on the shipper from that a customer purchases. These issues should be taken into consideration in the future research.

References

1) Holguin-Veras, J. and Thorson, E.: Trip Length Distributions in Commodity-Based and Trip-Based Freight Demand Modeling, Transportation Research Record 1707, pp. 37-48, 2000.
2) Sivakumar, A. and Bhat, C.: Fractional Split-Distribution Model for Statewide Commodity-Flow Analysis, Transportation Research Record 1790, pp. 80-88, 2002.
3) McMillen, P.: Probit with Spatial Autocorrelation, Journal of Regional Science, Vol. 32, No. 3, pp. 335-348, 1992.
4) Bhat, C.R. and Guo, J.: A Mixed Spatially Correlated Logit Model: Formulation and Application to Location Choice Modeling, Transportation Research Part B, Vol. 38, pp. 147-168, 2004.
5) Miyamoto, K., et al: Discrete Choice Model with Structuralized Spatial Effects for Location Analysis,

Transportation Research Record 1898, pp. 183-190, 2004.

6) Mohammadian, A., Haider, M., and Kanaroglou, P.S.: Incorporating Spatial Dependencies in Random Parameter Discrete Choice Models, Proceeding of the 84th Annual Transportation Research Board, 2005.

7) Bradlow, E.T., *et al*: Spatial Models in Marketing, The Spatial Models in Marketing Seminar at the 6th Invitational Choice Symposium, June 2004.

8) Wisetjindawat, W., Sano, S., and Matsumoto, S.: Commodity Distribution Model Incorporating Spatial Interactions for Urban Freight Movement, Proceeding of the 85th Annual Meeting of the Transportation Research Board, 2006.

Effects of Spatial Correlation on the Choice Behavior of Urban Goods Distribution*

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

This paper proposes a commodity distribution model that views commodity flows as the result of interplay between the shipper and customer of commodities. The commodity flows are originated from the customer who makes decision on purchasing a commodity from each of the suppliers. The model utilized a discrete choice framework accommodating the spatial interactions among customers in choice decision. The model is applied to the urban goods movement in the Tokyo Metropolitan Area. The results show significant improvement of the model that considers the spatial effects which confirms the assumption that the decision of customers is influenced by the others.

都市内物流における空間相関を考慮した分布貨物量推定モデルの構築*

Wisinee WISETJINDAWAT**・佐野可寸志***・松本昌二****

本研究では，個々の顧客の荷主選択行動を集計することにより，都市内の物資流動の分布貨物量を推定するモデルを提案する．本モデルは非集計選択モデルの枠組み用いているが，荷主の企業規模等の荷主属性や，顧客と荷主間の距離や，顧客周辺の同業者が当該荷主を選択する確率等を考慮して，顧客は荷主を選択すると仮定してモデル化を行う．東京都市圏物資流動調査のデータを用いてモデルの推計を行った結果，顧客間の相互作用を考慮したモデルは高い推計精度を持つことが確認され，これは顧客の荷主を選択するという意志決定は他の顧客の影響を受けているという仮説を支持するものであると考えられる．