

Commodity Distribution Model Incorporating Spatial Interactions for Urban Freight Movement

Wisinee Wisetjindawat
Department of Civil and Environmental Engineering
Nagaoka University of Technology
1603-1 Kamitomioka-Machi, Nagaoka, Niigata, 940-2188, Japan
Phone: +81-2-58-47-1611 (Ext. 6635)
Fax: +81-2-58-47-9650
Email: wisinee@stn.nagaokaut.ac.jp

Kazushi Sano
Department of Civil and Environmental Engineering
Nagaoka University of Technology
1603-1 Kamitomioka-Machi, Nagaoka, Niigata, 940-2188, Japan
Phone: +81-2-58-47-9616
Fax: +81-2-58-47-9650
Email: sano@nagaokaut.ac.jp

Shoji Matsumoto
Department of Civil and Environmental Engineering
Nagaoka University of Technology
1603-1 Kamitomioka-Machi, Nagaoka, Niigata, 940-2188, Japan
Phone: +81-2-58-47-9615
Fax: +81-2-58-47-9650
Email: shoji@nagaokaut.ac.jp

Paper Number: 06-1856

First Submission Date: 31st July 2005

Re-Submission Date: 14th November 2005

Final Submission Date: 31st March 2006

Word Count: Text (6,245) + Figures/Tables (5*250) = 7,495

*85th Annual Meeting of the Transportation Research Board
January 22-26, 2006. Washington D.C.*

ABSTRACT

The commodity distribution model proposed in this paper is developed in such a way that the movement of commodities is explained as an outcome of its flow through several freight agents in a supply chain. As commodity flow is fundamentally determined by the demand, the proposed model is developed using a discrete choice model considering the individual behavior of a customer to decide the suppliers from which to purchase and the amount of commodity acquired from each of the suppliers. The model not only takes into account the interplay between shipper and customer in a supply chain but also captures the spatial interactions among alternatives and among customers since spatial effects generally impact customer preference. In this study, we developed and compared several model specifications, that is, with and without incorporating spatial interactions. The empirical results of the model applied to analyze the urban commodity distribution in the Tokyo Metropolitan Area indicate that integrating both spatial interactions among alternatives and among customers statistically improves the model performance.

Keywords: Commodity Distribution, Mixed Logit, Spatial Effects, Urban Freight Movement

INTRODUCTION

Decision-making on a freight transport system is a challenge issue since the freight system causes a significant amount of energy consumption and pollution. Transport planners need a proper freight transport model in order to predict the effect of decision-making on a freight transport system. However, modeling of a freight system is a very difficult task because of its complexity since freight movement involves complicated linkages among many freight agents interacting in supply chains. The difficulty also is due to the heterogeneity of the freight system, whereby the characteristics of commodities vary in volume, weight, value, and shape and causes modeling of freight movement to be more difficult. In addition, current innovations in freight movement, such as Just-In-Time and third-party logistics, adopted in several freight organizations are necessary to consider while modeling freight systems. The microscopic modeling approach, which is currently gaining more favor in both passenger and freight demand modeling, is found to be suitable for modeling freight systems since it can reflect the real mechanism of movement. We attempt to develop a model that considers the individual behavior of each freight agent and its interaction with other freight agents in a supply chain.

Freight Transport Demand Modeling

Most models on freight movement heretofore rely on the traditional four-step approach in which the basic processes are freight generation, distribution, modal split, and freight traffic assignment. Four-step-based models generally can be categorized into two major groups: vehicle-trip-based and commodity-based approaches (1). The vehicle-trip-based approach has a similar process to the four-step approach of passenger trips in that it directly estimates generation and attraction freight trips from size indicators for trip generation (such as area, population, and number of firms). However, this approach neglects the fundamental concept that the movement of vehicles is the result of the intention to move commodities. As freight trips are directly generated, it is therefore difficult to evaluate policies related to the change in the characteristics of commodity movement. The commodity-based approach, on the other hand, generates the amounts of commodity production and consumption instead of the number of vehicle trips, resulting in a more realistic and policy-sensitive model. Holguin-veras and Thorson (1) elaborately described the differences, advantages, and disadvantages of the approaches.

Earlier freight distribution models of the four-step approach have focused on the gravity model for modeling the distribution pattern of freight movements (2). Gravity models calculate the flows (either in term of number of trips or tons of commodities) between each pair of origin and destination as a function of generation and attraction of the origin and destination zones weighted by an impedance term that represents transport costs between the zone pair. Ashtakala and Murthy (3) developed a set of gravity models for several types of commodities for province-wide freight movement in Alberta based on impedance such as distance and travel time. Another example of a gravity model for freight distribution is SMILE (4) in which the commodity distribution is based on the comparative price differences and the impedance of geographical and organizational differences. However, the most important disadvantage of the gravity models is that they lack the fundamentals of freight movement.

Recently, Sivakumar and Bhat (5) proposed a new approach, which is the fractional split distribution model, for modeling inter-regional commodity flows. The approach adopted a logit model to estimate the fraction of freight consumed at the destination zone that originated from each of the production zones. This idea conforms to the basis of freight

movement, that is, freight movement is generated by the demand of commodities at the destination that is met by the flow of commodities from one or more points of origin (5). The empirical application confirmed that the fractional split distribution model provides better results compared with those of the gravity model. Nevertheless, there are some limitations; since this model deals with freight movement at a zone level; it is difficult to consider the issues considering the behavior of freight agents.

Freight Agent Interactions

The difficulty of modeling freight transport is caused by the complicated relationships among freight agents, including shippers, customers, carriers, and administrators (2). Each freight agent plays a different role in freight movement and interacts with the others while trying to move commodities through supply chains. Such interactions are the key characteristic and must consider in modeling freight transport. For several years, there have been many attempts to model the interactions among freight agents. For the interplay between shipper and carrier, McFadden et al. (6) proposed a model for estimating the freight transport decision joint choice between mode and shipment size based on the concept of the optimization of inventory cost. However, the estimation using the model is very complicated and difficult to apply in practice. Instead of the inventory-based model, Abdelwahab and Sargious (7) utilized a discrete choice model for modeling the mode choice and shipment size of freight movement. As shipment is, in fact, a continuous variable, Holguin-Veras (8) overcame the limitation of the above model by formulating a mode choice and shipment size model using discrete-continuous choice formulation and by treating the shipment size as a continuous variable. However, the above models consider only the interaction between two freight agents. Recently, Boerkamps (9, 10) incorporated the interactions among all the freight agents in a supply chain into modeling of urban freight movement.

Interplay among freight agents also can be view as interactions in a space. In urban economics, related firms that gather together will gain benefits such as lower cost of production and a greater market share that one firm can achieve. A firm, while making a choice decision, therefore, does not act in isolation; in contrast, Firms are influenced by others who located nearby. Models considering spatial dependences are based on the belief that decision maker's preferences are correlated and take into account the effect of heterogeneity on the consumer's preference. Bradlow et al. (11) provided a review of various spatial models in marketing science. In modeling, spatial interaction currently trends to the use of the mixed logit model since the model allows several specifications and has a simple estimation technique. Train (12) and Walker (13, 14) provide summaries on theory, specifications, and estimations of various types of mixed logit models. The earlier work on spatial effects in the discrete choice model is the work of McMillen (15) in which spatial autocorrelation was incorporated into a probit model. Bhat and Guo (16) compared several specifications of discrete choice models. They found that the mixed multinomial logit model with spatial correlation statistically fits the data better than the other discrete choice models. Mohammadian et al. (17) integrated spatial dependences in the deterministic part and considered the preference heterogeneity in the error part of the mixed logit model, which significantly improved the model performance. In the works of (15), (16), and (17), spatial effects were incorporated into residential choice. However, spatial effects have never been applied to purchasing choice decisions in freight transport which is the objective of our work.

Objective of Our Study

The commodity distribution model proposed here is a commodity-based model incorporating supply chain characteristics and spatial effects. In this work, we extend the concept of the fractional split distribution model and apply it to urban freight movement. Urban freight movement in particular can be viewed as the outcome of the choice decisions of each individual customer on purchasing commodities originating from each supplier in a supply chain. In addition, the effects of spatial interactions among alternative zones and among customers are integrated into the choice decision process. We utilize a mixed logit framework with the specifications of the deterministic and disturbance terms to consider the spatial interactions. The model covers the first two steps of the traditional four-step approach in which commodity generation and distribution are undertaken.

This paper is structured into four sections. The next section provides the proposed model structure, concept, mathematical formulations, and estimation method. The empirical results and discussions on the proposed model applied to urban freight movement in the Tokyo Metropolitan Area are subsequently presented. The final section is the conclusion and recommendations on the proposed model.

MODEL SPECIFICATIONS

Model Structure

The proposed commodity distribution model is an application of the spatial discrete choice model to the purchasing behavior of each individual firm. In this paper, a firm is defined as a shipper or a customer when the firm generates or attracts a commodity, respectively. Based on the fundamental concept that consumption demand governs commodity flows, the model makes a connection from demand to supply by assuming a customer as a decision-maker who selects to purchase commodities from available shippers according to the attractiveness of the supplier, the characteristics of the customer, and the structure of the supply chain of the commodity. The spatial dependences that the spatial location of shippers and customers affects the decision process are also incorporated. The model consists of two sub models: commodity generation and commodity distribution.

Commodity Generation

The amounts of commodity production and consumption are used to represent the total amounts of commodities produced and consumed by each firm. Regression models to estimate the amounts of commodity production and consumption are developed using firm size indicators such as number of employees and floor area.

$$G_i^k = \alpha_1^k x_{i1} + \alpha_2^k x_{i2} + \dots + \alpha_n^k x_{in}, \quad G_I^k = \sum_{i \in I} G_i^k \quad (1)$$

$$A_j^k = \chi_1^k y_{j1} + \chi_2^k y_{j2} + \dots + \chi_m^k y_{jm}, \quad A_J^k = \sum_{j \in J} A_j^k \quad (2)$$

Where, G_i^k is the amount of commodity k produced by shipper i . α_n is a parameter for production indicator n of commodity k . x_{in} is the production indicator n of firm i . G_I^k is the total amount of commodity k generated in zone I . A_j^k is the amount of commodity k consumed by firm j . χ_n is a parameter for consumption indicator m of commodity k . y_{jm}

is the consumption indicator m of firm j . A_j^k is the total amount of commodity k attracted into zone J .

Commodity Distribution

The supply chain structure is the key characteristic of freight flows. The distribution channel, defined as the path connecting customers and shippers of a commodity, describes the ways that commodities are distributed through each linkage of the supply chain (9, 10). In a common supply chain, as shown in Figure 1, there are a number of possible distribution channels, depicted as the lines connecting each pair of agents. For example, a retailer may purchase a commodity from a wholesaler or a higher level.

The commodity distribution derived from the commodity flow between shipper and customer is calculated by multiplying the fraction of a customer purchasing a commodity from a shipper by the total consumption amount of the customer as follows:

$$Q_{ij}^k = P_j^k(i) \cdot A_j^k \quad (3)$$

where Q_{ij}^k is the commodity flow between shipper i to customer j for commodity k . $P_j^k(i)$ is the fraction of commodity k that customer j purchases from shipper i . A_j^k is the total consumption amount of customer j for commodity k .

The supply chain structure of the commodity is incorporated into the proposed model. $P_j^k(i)$ consists of three parts: 1) distribution channel probability, 2) zone choice probability, and 3) shipper choice probability. The probability is used instead of the fraction of consumption since a firm will not purchase commodities from all available suppliers. The probability will then be converted to the fraction of purchases from each of the selected suppliers once the suppliers being selected by each customer are identified. The conversion is a straightforward weighting of the purchasing fraction of each of the selected suppliers according to the calculated probabilities. The product of all three parts yields the probability of a given shipper being selected. The following shows the mathematical form of the model:

$$P_j^k(i) = P_j(C^k) \cdot P_j(z|C^k) \cdot P_j(i|C^k, z), \quad i \in C^k, z \quad (4)$$

where $P_j(C^k)$ 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 by customer j , given distribution channel C and zone z .

First, the probability of a distribution channel being selected for purchasing a commodity, $P_j(C^k)$, is determined from the empirical data. The survey data provides the percentage of each distribution channel being used to distribute each type of commodity. Second, the probability of a zone being selected, $P_j(z|C^k)$, is a location choice selection problem with multiple alternative zones. We utilize a spatial mixed logit model and assume a utility function for zonal attractiveness represented by a firm's spatial distribution and attractiveness indicators, such as the number of firms and total amount of commodity generated within a zone. The deterministic and disturbance terms of the mixed logit model are specified to incorporate the spatial interactions among decision makers and among alternative

zones, respectively. A customer is assumed to select the zone with the maximum utility among zone alternatives. The details on this part are provided in the remaining sections.

Lastly, the probability of a shipper being selected, $P_j(i|C^k, z)$, is given by identifying the shippers from which a customer purchases. However, due to the limitation of the survey data that is not identified the exact shipper from which each customer makes purchases, we assume that the shipper probability can be derived directly from the exponential of the production amount of a shipper divided by the sum of the exponential of all shippers in the given distribution channel and zone, as shown by Equation (5). The conditional probability that shipper i is chosen by customer j , given C^k and z , is

$$P_j(i|C^k, z) = \frac{\exp(G_i^k)}{\sum_{i \in C_j} \exp(G_i^k)} \quad (5)$$

where G_i^k is the amount of commodity k produced by shipper i . i' is a shipper in C_j , which is the set of shippers in the given C^k and z . Once the shipper probabilities from the above three parts are determined, random numbers are generated and used with the shipper probabilities to indicate the selected shippers for each customer. The fraction of commodity purchased from each selected shipper is determined directly by weighting from the calculated probability. Multiplication of the fraction and the total amount of commodity consumption of the customer yields the commodity flow from shipper to customer. Summing all commodity flows among firms of each zone finally yields the commodity OD matrix.

Zone Choice Probability with Spatial Interactions, $P_j(z|C^k)$

Spatial dependence among people preferences plays a key role in decision choice behavior. 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.

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

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.

A mixed logit model, which consists of a flexible probit-like term and an additive iid extreme value term, is utilized to explain the choice decision process and incorporates the spatial interaction in the disturbance term. Based on the work of Ben-Akiva et al. (18), the vector of the utility of alternatives for customer j takes the general form of the mixed logit model as follows:

$$\mathbf{U}_j = \mathbf{X}_j \boldsymbol{\beta} + \mathbf{F}_j \mathbf{T} \boldsymbol{\zeta}_j + \mathbf{v}_j. \quad (7)$$

The deterministic part of the utility function is represented by the first term ($\mathbf{X}_j \boldsymbol{\beta}$), where $\boldsymbol{\beta}$ is a $[K \times 1]$ vector of unknown parameters and \mathbf{X}_j is a $[Z \times K]$ matrix of explanatory variables for customer j . The disturbance term is depicted by the second and third terms ($\mathbf{F}_j \mathbf{T} \boldsymbol{\zeta}_j$ and \mathbf{v}_j). $\mathbf{F}_j \mathbf{T} \boldsymbol{\zeta}_j$ represents the covariance structure among alternative zones, where \mathbf{F}_j is a $[Z \times Z]$ matrix of factor loading that is used to construct the pattern of correlation, \mathbf{T} is a $[Z \times Z]$ matrix of standard deviation with σ on the diagonal, and $\boldsymbol{\zeta}_j$ is a $[Z \times 1]$ vector of random numbers with zero mean and unit variance; in this study, we assume $N(0,1)$. \mathbf{v}_j is a $[Z \times 1]$ vector of iid gumbel random variables.

The probability of alternative z being selected by customer j is the integral of the standard logit probability over the density of parameter. The conditional probability that zone z is chosen by customer j , given distribution channel C^k , is formulated as

$$P_j(z|C^k) = P_{zj} = \int_{\boldsymbol{\zeta}} \Lambda(z|\boldsymbol{\zeta}) n(\boldsymbol{\zeta}, I_Z) d\boldsymbol{\zeta} = \int_{\boldsymbol{\zeta}} \left(\frac{\exp(X_{zj} \boldsymbol{\beta} + F_{zj} \mathbf{T} \boldsymbol{\zeta}_j)}{\sum_{z' \in Z} \exp(X_{z'j} \boldsymbol{\beta} + F_{z'j} \mathbf{T} \boldsymbol{\zeta}_j)} \right) n(\boldsymbol{\zeta}, I_Z) d\boldsymbol{\zeta} \quad (8)$$

where X_{zj} is z^{th} row of matrix \mathbf{X}_j representing attributes of zones for customer j . $\boldsymbol{\beta}$ is a matrix of parameters to be calibrated. The disturbance term $F_{zj} \mathbf{T} \boldsymbol{\zeta}_j$ depicts the covariance structure for customer j . F_{zj} is z^{th} row of the matrix in \mathbf{F}_j . z' is a zone in Z , which is a set of all alternative zones. $\Lambda(z|\boldsymbol{\zeta})$ is the logit part of the model, and is the probability that the choice is z given $\boldsymbol{\zeta}$. $n(\boldsymbol{\zeta}, I_Z)$ is the joint density function of $\boldsymbol{\zeta}$, and is a product of the standard univariate normal.

$$n(\boldsymbol{\zeta}, I_Z) = \prod_{z=1}^Z \Phi(\zeta_z) \quad (9)$$

Incorporating the Correlation among Alternatives in the Error Term

The correlations among alternatives can be taken into account by specifying the disturbance term. The spatial autoregressive term inserted into the disturbance term of the probit model is successfully applied to housing choice behavior in the model of McMillen (15). Ben-Akiva et al. (18) suggested to add the generalized autoregressive term to the disturbance term of mixed logit. We therefore utilize the generalized autoregressive term to explain the spatial correlation among alternatives, defined as follows:

$$\boldsymbol{\xi}_j = \rho \mathbf{W} \boldsymbol{\xi}_j + \mathbf{T} \boldsymbol{\zeta}_j = (\mathbf{I} - \rho \mathbf{W})^{-1} \mathbf{T} \boldsymbol{\zeta}_j \quad \text{or} \quad \mathbf{F}_j = (\mathbf{I} - \rho \mathbf{W})^{-1}, \quad (10)$$

Where, $\boldsymbol{\xi}_j$ is the disturbance term of mixed logit. ρ is a scalar unknown parameter. \mathbf{I} is a $[Z \times Z]$ identity matrix. \mathbf{W} is a $[Z \times Z]$ weight matrix identifying the correlation among alternative zones. \mathbf{F}_j in this study is assumed to be the same for all customers.

Specification of the weight matrix, which indicates the similarity between zones, is important to explain the spatial interaction pattern of the model. There are several ways to specify the weight matrix [see (19) for several specifications of the spatial weight matrix]. Typically, the adjacency weight matrix and distance-based weight matrix are used to measure the level of correlation between zones. Adjacency weight matrix, as described in (15) and (16), is a symmetric matrix with zeros on the diagonal element, and the off-diagonal element (w_{IJ}) of the adjacency weight matrix is defined as

$$w_{IJ} = \frac{c_{IJ}}{\sum_J c_{IJ}}, \quad (11)$$

where c_{IJ} equals 1 if zone J shares a common boundary with zone I and 0 otherwise.

The distance-based weight matrix is often used to specify the spatial weight matrix. We utilize an inverse distance weighting function,

$$w_{IJ} = \frac{1}{d_{IJ}^\gamma}, \quad (12)$$

where w_{IJ} is an element of weight matrix \mathbf{W} . d_{IJ} is the distance between zones I and J . γ is an unknown parameter. Substituting ξ_n into Equation (7), the logit part of the probability in Equation (8) becomes

$$\Lambda(z|\zeta_j) = \frac{\exp[X_{zj}\boldsymbol{\beta} + (\mathbf{I} - \rho\mathbf{W})_z^{-1}\mathbf{T}\zeta_j]}{\sum_{z \in Z} \exp[X_{zj}\boldsymbol{\beta} + (\mathbf{I} - \rho\mathbf{W})_z^{-1}\mathbf{T}\zeta_j]}, \quad (13)$$

where $(\mathbf{I} - \rho\mathbf{W})_z^{-1}$ is z^{th} row of the matrix $(\mathbf{I} - \rho\mathbf{W})^{-1}$.

Incorporating the Correlations among Customers in the Deterministic Term

Spatial dependences do not only appear in the correlations among alternative zones but also include the correlations among customers since a consumer's preference is also influenced by the decision of other consumers. Mohammadian et al. (17) incorporated spatial dependences into a mixed logit model to explain housing choice behavior. The interactions among decision makers are accommodated by adding the interaction part into the fixed variable part of the spatial mixed logit model. The application of the model shows an interesting improvement of the results with the spatial interaction over those without the spatial interaction. We apply this concept to explain the interaction among customers. Two specifications of distance decay functions are utilized: negative exponential and inverse distance. For each element of the deterministic part (V_{zj}), the model after adding the interactions becomes

$$V_{zj} = X_{zj}\boldsymbol{\beta} + \phi_{zj} = \sum_{k=1}^K \beta_{zk} x_{zjk} + \lambda \sum_{s=1}^S y_{zs} \exp(-d_{js}^\delta) \quad (14)$$

$$V_{zj} = X_{zj}\boldsymbol{\beta} + \phi_{zj} = \sum_{k=1}^K \beta_{zk} x_{zjk} + \lambda \sum_{s=1}^S y_{zs} \frac{1}{d_{js}^\delta}, \quad (15)$$

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 . δ is a scalar unknown parameter. Substituting Equation (14) or (15) into Equation (13), the logit part yields

$$\Lambda(z|\zeta_j) = \frac{\exp[X_{zj}\boldsymbol{\beta} + \phi_{zj} + (\mathbf{I} - \rho\mathbf{W})_z^{-1}\mathbf{T}\zeta_j]}{\sum_{z' \in Z} \exp[X_{z'j}\boldsymbol{\beta} + \phi_{z'j} + (\mathbf{I} - \rho\mathbf{W})_{z'}^{-1}\mathbf{T}\zeta_j]}, \quad (16)$$

where ϕ_{zj} is the correlation between customer j and other customers for alternative zone z , which is presented by Equations (14) and (15).

Estimation of the Model

In the proposed model, input, output, and parameters to be calibrated in each of the models, as summarized in Table 1, are needed. For the zone choice model which utilizes spatial mix logit, we must estimate the vector of scalar unknown parameters $\boldsymbol{\beta}$ and parameter σ in the error structure for the model in Equation (13). Likewise, the model need to estimated the vector of scalar unknown parameters $\boldsymbol{\beta}$, the scalar parameters λ and δ , and the parameter σ in error structure for the model described by Equation (16). Let us denotes $\boldsymbol{\theta}$ as the vector of joint parameters of all parameters to be estimated. Since the integral form in Equation (8) cannot be calculated, the probability is approximated through the simulation of any given value of ζ_j^r , where r denotes draw of j from the distribution of ζ . The average value of these probabilities yields the following simulated probability.

$$\hat{P}_j(z|\boldsymbol{\theta}) = \frac{1}{R} \sum_{r=1}^R \Lambda(z|\boldsymbol{\theta}, \zeta_j^r) \quad (17)$$

where R is the total number of draws.

Incorporating Equations (13) and (16) into Equation (17), we obtain the simulated probabilities of the model incorporating spatial correlation among zone alternatives and the model incorporating spatial correlations among both zone alternatives and customers as Equations (18) and (19), respectively.

$$\hat{P}_j(z|\boldsymbol{\theta}) = \frac{1}{R} \sum_{r=1}^R \frac{\exp[X_{zj}\boldsymbol{\beta} + (\mathbf{I} - \rho\mathbf{W})_z^{-1}\mathbf{T}\zeta_j^r]}{\sum_{z' \in Z} \exp[X_{z'j}\boldsymbol{\beta} + (\mathbf{I} - \rho\mathbf{W})_{z'}^{-1}\mathbf{T}\zeta_j^r]} \quad (18)$$

$$\hat{P}_j(z|\boldsymbol{\theta}) = \frac{1}{R} \sum_{r=1}^R \frac{\exp[X_{zj}\boldsymbol{\beta} + \phi_{zj} + (\mathbf{I} - \rho\mathbf{W})_z^{-1}\mathbf{T}\zeta_j^r]}{\sum_{z' \in Z} \exp[X_{z'j}\boldsymbol{\beta} + \phi_{z'j} + (\mathbf{I} - \rho\mathbf{W})_{z'}^{-1}\mathbf{T}\zeta_j^r]} \quad (19)$$

The parameters can be estimated by the maximum likelihood method. Since the true log-likelihood cannot be calculated, the simulated maximum likelihood technique is used and the formulation can be written as

$$SLL = \sum_{j=1}^N \sum_{z=1}^Z y_{zj} \ln \hat{P}_j(z|\boldsymbol{\theta}), \quad (20)$$

where N is the total number of customers.

The simulation steps are 1) draw a value ζ from a distribution in which, in this study, $N(0,1)$ gives the best results compared with the other distributions. Then, label ζ_j^r for draw number r of customer j ; 2) calculate the logit part $\Lambda(z|\boldsymbol{\theta}, \zeta_j^r)$ using ζ_j^r , and 3) repeat steps 1) and 2) as many times as the desired number of draws.

The number of draws should be large enough to avoid bias. Generally, pseudorandom draw or Halton draw are used for the simulation. Halton draw is found to be more efficient than pseudorandom draw when the number of dimensions is small. Otherwise, there will be a large bias (13). However, because the number of dimensions in this study is quite large, we utilize pseudorandom draw with 500 draws, which yields a stable estimation for our simulation.

EMPIRICAL CASE STUDY

Data Sets

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 Ibaraki. In this study, we adopt the large zoning system for calculation purposes. The area is divided into 56 zones comprising 52 zones within the study area and 4 zones in the prefectures near the study area for the analysis of the external trips. The area of each zone ranges from 100 to 500 km².

The proper classification of commodity types and industry types is a necessary step prior to model calibration. We categorize commodities into the following 8 types.

- 1) Agricultural Products
- 2) Forestry Products
- 3) Mineral Products
- 4) Metal and Machinery Products
- 5) Chemical Products
- 6) Light Industry Products
- 7) Other Products
- 8) Wastes and Scraps

In the same way, industry types are classified into the following 13 groups.

- 1) Agriculture, Forestry, and Fishery
- 2) Mining
- 3) Construction
- 4) Chemical Manufacturers
- 5) Metal Manufacturers
- 6) Machinery Manufacturers
- 7) Other Manufacturers
- 8) Material Wholesalers

- 9) Product Wholesalers
- 10) Retailers
- 11) Warehouses
- 12) Electricity, Gas and Water Suppliers
- 13) Service and Government Work

Model calibration is performed using the database of the Tokyo Metropolitan Goods Movement Survey (TMGMS) collected from firms throughout the TMA by the Ministry of Land, Infrastructure and Transport. We utilize the data of 1982, which is the most complete database. The data consists of records on the movements of commodities and trucks by each firm and was collected from approximately 46,000 firms, corresponding to three percent of all firms in the study area. Each record provides information regarding firm characteristics, commodity movement, and truck movement (including, industry type, location, number of employees, commodity type and weight carried in and out, delivery frequency, locations of origin and destination of each freight trip, truck size, carrier type, and other related information).

Results and Discussions

Light industry products (including processed food products, papers, threads, fabrics, and other related products) are selected for analysis in this study, because it is the main commodity of urban freight movement. The number of truck trips of this commodity type is the largest and accounts for 30 percent of the total number of trips of all commodity types. The probability of a shipper being selected consists of three parts: distribution channel choice probability ($P_j(C^k)$), zone choice probability ($P_j(z|C^k)$), and shipper choice probability ($P_j(i|C^k, z)$). First, $P_j(C^k)$ is determined directly from the survey data by the percentage of commodity weight being purchased, as shown in Table 2. Second, a model to determine $P_j(z|C^k)$ is developed to consider spatial interactions. The details of the model are discussed in the next paragraph. Lastly, $P_j(i|C^k, z)$ is calculated directly from the variables indicating firm size, which, in this case, is the total amount of commodity production, since the survey data is limited and does not identify the exact shipper from which a firm makes purchases. At present, this model cannot be calibrated but will be considered in the future improvement.

A model for $P_j(z|C^k)$ is developed separately for each commodity type that is purchased by each industry type of customer from each industry type of shipper. Three models for three distribution channels mainly used for the distribution of light industry products are selected for discussion in this paper.

- 1) DC1: distribution channel for service and government work (*Ic13*) purchasing light industry products (*k6*) from retailers (*Is10*)
- 2) DC2: distribution channel for retailers (*Ic10*) purchasing light industry products (*k6*) from wholesalers (*Is9*)
- 3) DC3: distribution channel for other manufacturers (*Ic7*) purchasing light industry products (*k6*) from other manufacturers (*Ic7*)

Model estimation is performed using GAUSS programming language. The independent variables considered in this analysis consist of three main types: zonal impedance variables, zonal attraction variables, and correlation variables. The average distance between zones obtained from the empirical data is used as the zonal impedance variable. The zonal attractiveness variables are the total amount of commodity type k produced in a zone (GEN), the number of establishments of industry type C^k (NC), population (POP), area, and the

number of employees (EMP). The spatial weight matrix represents the spatial relationships among alternative zones. Two types of specification of the weight matrix are used in this study. In choosing between the adjacency weight matrix and the distance-based weight matrix, the one that gives the better result is selected to explain the correlation pattern for each model.

The estimated results from three modeling types DC1, DC2, and DC3 are compared in Table 3. MNL is a common multinomial logit model without considering any kind of spatial interactions. SML-A is the model that incorporates only the spatial interaction among zone alternatives while SML-AD includes spatial interactions both among zone alternatives and among customers. For model DC1, the weight matrix used in SML-A and SML-AD is an adjacency weight matrix specified by Equation (11). The interaction among decision makers takes the negative exponential form, as presented in Equation (14). The t-statistics of zonal attractiveness variables are statistically significant for all 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 zonal attractiveness variables implies that a firm is likely to purchase a larger quantity from the zones having the larger area and the greater number of establishments. The negative signs of the distance variables of all models reveal that the farther the zone is away from the customer, the less the quantity of commodities that will be purchased from there. The spatial parameters are statistically significant for both SML-A and SML-AD, confirming that interactions both among zone alternatives and among customers have an influence on the decision-making process. We apply both the adjusted log-likelihood ratio and Akaike Information Criterion (AIC) to test model performance. The AIC test is utilized in order to penalize log-likelihood improvements due to a large number of parameters. A higher value of the adjusted log-likelihood ratio is preferable, while a lower value of AIC is more desirable. Considering the results of the two tests performing on the three models, SML-AD has a significantly improved model performance, implying that the interactions with other customers have greater effects compared with the interactions among zone alternatives. In addition, the validation results shown in graphs a1), a2), and a3) of Figure 2 confirm the SML-AD gives the best fit in the prediction of consumption choice.

The models of purchasing zone choice for DC2, the distance-based weight matrix specified as Equations (12) and (15) is suitable for modeling the interaction among zone alternatives and the interaction among customers, respectively. All the zonal attractiveness variables of all models are statistically significant in terms of the t-statistic with reasonable signs. A positive value of the zonal attractiveness variables implies that the larger purchasing quantity comes from the zones having the higher zonal attractiveness. On the contrary, a negative value of the impedance variable indicates that the farther zones are less attractive to customers for purchasing the commodity. The spatial correlations are also significant for correlations both among zone alternatives and among decision makers. The negative sign of ρ means that increasing the purchasing quantity from a zone leads to a decrease in the purchasing quantity from other zones. Considering the results of the adjusted log-likelihood ratio and AIC test, SML-AD has the best model performance, implying that, similarly to the previous model, the interaction among customers greatly affects their choice decision. Graphs b1), b2), and b3) of Figure 2 show the validation results for the three modeling types. SML-AD gives the best fit compared with the other two modeling types.

For model DC3, the weight matrix used for this distribution channel is the distance-based weight matrix shown by Equations (12) and (15) for modeling the interaction among zone alternatives and the interaction among customers, respectively. Similarly, the zonal attractiveness variables and impedance variables of all models are statistically significant with reasonable signs. The variables representing spatial correlations are statistically significant, implying the variables are meaningful for explaining the purchasing choice behavior. Similarly, the best model performance of the SML-AD model in terms of the adjusted log-

likelihood ratio and AIC test and the validation results in graphs c1), c2), and c3) of Figure 2 confirm that the customer is statistically influenced by the spatial interactions among zone alternatives and with other customers.

Comparing the results for the models of the three distribution channels, the number of establishments for DC1 is the highest compared with those of the other distribution channels. This means that the number of establishments in a zone more strongly affects the purchasing quantity of end users from retailers compared with those of other industry types. The amount of generated commodities for DC3 has the greater value than that for DC2, while the population for DC2 is greater than that for DC3, implying that the amount of commodity generation has a greater affect when commodities are distributed between manufacturers than between the lower levels of the supply chain. For commodity distribution between wholesalers and retailers, on the other hand, the population in a supply zone is more important. The impedance variable for DC3 has the highest value, whereas the variable for DC1 has the lowest value. The results emphasize the fact that commodities are usually distributed between manufacturers at longer distances than those distributed between wholesalers and retailer and between retailers and end users, respectively. The correlation variables for DC1, which uses an adjacency weight matrix, has different structures from those of the other upper levels of the supply chain, indicating that only zones that share the zone boundary will be impacted when the end user changes the consumption quantity of the commodities purchased from retailers. For DC2 and DC3, which have the similar structures, the larger correlation variable of ρ of DC2 than that of DC3 reveals that when customers increase purchasing quantity from a zone, the consumption quantity from the nearer zones for DC3 will decrease more than that for DC2. This can be interpreted to mean that manufacturers do not always change the supplier's manufacturers since it is true that the commodities flowing between manufacturers are mostly distributed within the same company.

CONCLUSION

We have proposed an alternative approach for modeling commodity flows in an urban area. The model takes into account the fundamentals of commodity movement, which is the outcome of commodities flowing through supply chains. Each individual consumption point is linked to a production point according to the attractiveness of the production point and their relationship in a supply chain. The model results in commodity flow from firm to firm over the entire area. The complex interactions among freight agents in a supply chain, as well as the spatial interactions affecting each agent's behavior, are incorporated. Using the proposed model, the freight movement pattern can be predicted due to the implementation of several logistics initiatives, not only in large-scale changes (such as a new transportation infrastructure), but also in small-scale changes (such as changes in the supply chain structure).

The model was applied to urban freight movement in the TMA and the results for three models, without spatial interaction, with interaction among alternative zones, and with both interactions among alternatives and among customers, were compared. The results indicate that the spatial interaction significantly affects the decision-making process. In particular, the interaction with other customers has a greater influence on a customer than the interaction among alternative zones.

The present implementation of the model still does not consider some aspects, for example, the effects of changes in the utilization of distribution channels since firms may have reorganize the supply chain structure to improve the efficiency. In addition, the model must overcome the limitation of survey data in which there is little information on the shipper from which a firm purchases. These issues remain to be considered in the future.

REFERENCES

1. Holguin-Veras, J. and Thorson, E. Trip Length Distributions in Commodity-Based and Trip-Based Freight Demand Modeling. *Transportation Research Record 1707*, 2000, pp. 37-48.
2. Taniguchi, E., Thompson, R.G., Yamada, T., and Duin, R.V. *City Logistics: Network Modeling and Intelligent Transport Systems*. Pergamon, Netherlands, 2001.
3. Ashtakala, B. and Murthy, A.S.N. Optimized Gravity Models for Commodity Transportation, *Journal of Transportation Engineering, ASCE, 114 (4)*, 1988, pp. 393-408.
4. Tavasszy, L.A. and Van Der Rest, H. Scenario-Wise Analysis of Transport and Logistics Systems with a SMILE, *Selected Proceedings of the 8th World Conference on Transportation Research*, 1999, pp.409-419.
5. Sivakumar, A. and Bhat, C. Fractional Split-Distribution Model for Statewide Commodity-Flow Analysis. *Transportation Research Record 1790*, 2002, pp. 80-88.
6. McFadden, D., Winston, C., and Boersch-Supan, A. Joint Estimation of Freight Transportation Decisions under Nonrandom Sampling. *Analytical Studies in Transport Economics*, A. Daugherty, ed., Cambridge University Press, New York.
7. Abdelwahab, W.M. and Sargious, M.A. A Simultaneous Decision-Making Approach to Model the Demand for Freight Transportation, *Canadian Journal of Civil Engineering*, Vol. 18 (3), 1985, pp. 515-520.
8. Holguin-Veras, J. Revealed Preference Analysis of Commercial Vehicle Choice Process, *Journal of Transportation Engineering (ASCE)*, Vol. 128, No. 4, 2002, pp. 336-346.
9. Boerkamps, J.H.K., Binsbergen, A.J.V., and Bovy, P.H.L. Modeling Behavioral Aspects of Urban Freight Movement in Supply Chains. *Transportation Research Record 1725*, 2000, pp. 17-25.
10. Boerkamps, J.H.K. and Binsbergen, A.J.V. GoodTrip – A New Approach for Modeling and Evaluation of Urban Goods Distribution. *Proceedings of the first International Conference on City Logistics: City Logistics I*, Australia, 1999, pp. 175-186.
11. Bradlow, E.T., Bronnenberg B., Russell G.J., Arora N., Bell D.R., Deepak S.D., Hofstede F., Sismeiro, C., Thomadsen, R., and Yang, S. Spatial Models in Marketing, *The Spatial Models in Marketing Seminar at the 6th Invitational Choice Symposium*, June 2004.
12. Train, K.E. *Discrete Choice Methods with Simulation*, Cambridge, United Kingdom, 2003.
13. Walker, J. Mixed Logit (or Logit Kernel) Model: Dispelling Misconceptions of Identification, *Transportation Research Record 1805*, 2002, pp. 86-98.
14. Walker, J. Extended Discrete Choice Models: Integrated Framework, Flexible Error Structures, and Latent Variables, *Ph.D. Dissertation*, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, 2001.
15. McMillen, D.P. Probit with Spatial Autocorrelation, *Journal of Regional Science*, Vol. 32 (3), 1992, pp. 335-348.
16. Bhat, C.R. and Guo, J. A Mixed Spatially Correlated Logit Model: Formulation and Application to Residential Choice Modeling, *Transportation Research Part B (38)*, 2004, pp. 147-168.
17. 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*. CD-ROM. Transportation Research Board, 2005.
18. Ben-Akiva, M., Bolduc, D., and Walker, J. Specification, Identifications, and Estimation of the Logit Kernel (or Continuous Mixed Logit) Model. *PDF (DRAFT)*, 2001.
19. Miaou, S. and Sui, D. Implications of Changing Demographic and Socioeconomic Structures on Highway Safety: A Texas Initiative, *Final Report*, 2004, pp. 19-21.

LISTS OF FIGURES

FIGURE 1. Distribution Channels of a Supply Chain

FIGURE 2. Estimated and Observed Fractions of Commodity Flows for Zone Choice Model of DC1, DC2, and DC3

LISTS OF TABLES

TABLE 1. Summary of Input, Output, and Parameters in each Model

TABLE 2. Percentage of Customer Industry Type I_c Purchases Commodity Type k from Shipper Industry Type I_s

TABLE 3. Estimated Results for DC1, DC2, and DC3

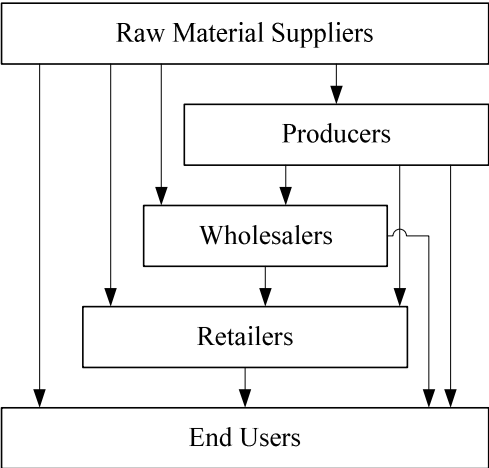
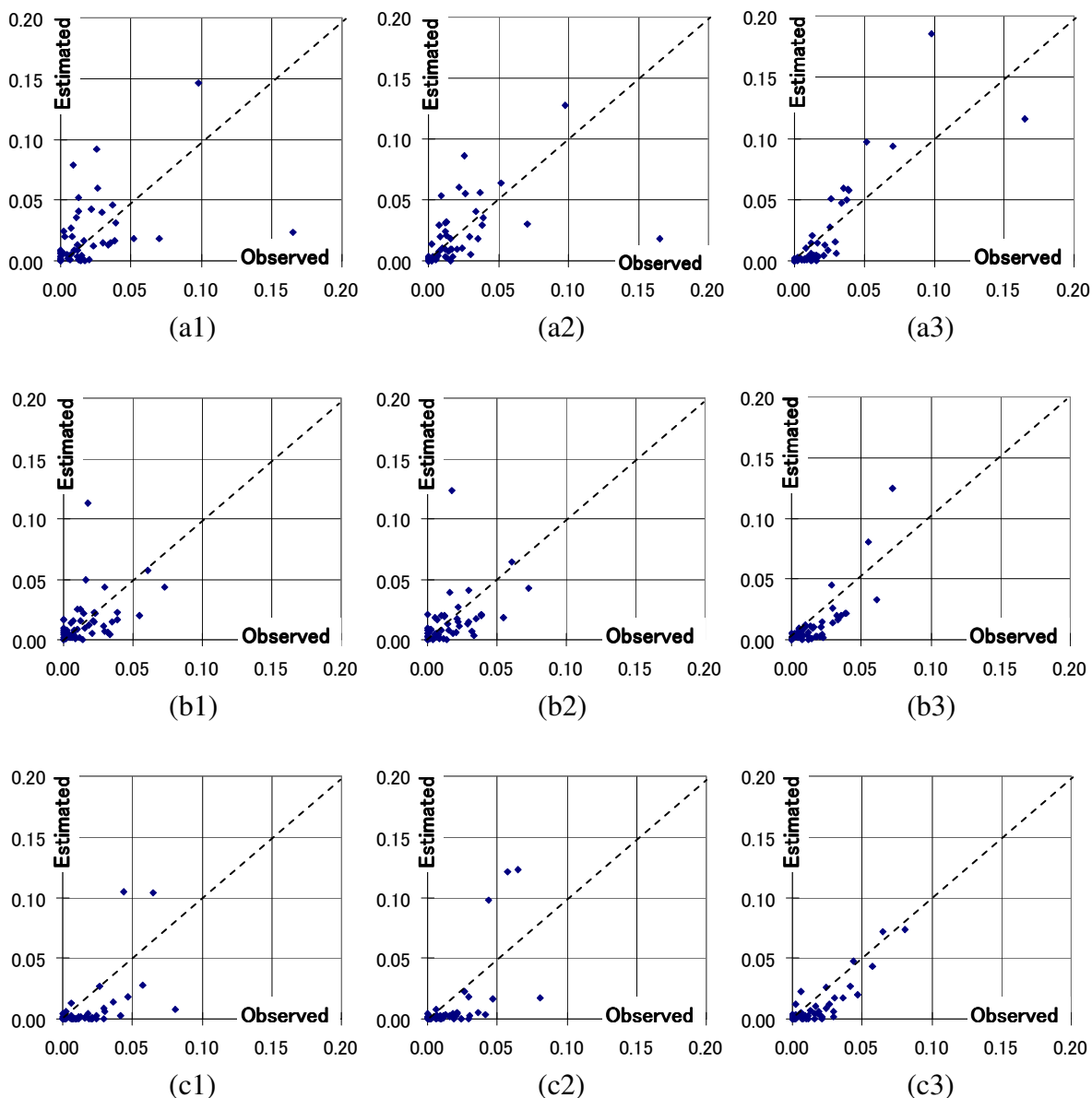


FIGURE 1. Distribution Channels of a Supply Chain



Note: (a1) Validation of MNL Model for DC1
 (a2) Validation of SML-A Model for DC1
 (a3) Validation of SML-AD Model for DC1

 (b1) Validation of MNL Model for DC2
 (b2) Validation of SML-A Model for DC2
 (b3) Validation of SML-AD Model for DC2

 (c1) Validation of MNL Model for DC3
 (c2) Validation of SML-A Model for DC3
 (c3) Validation of SML-AD Model for DC3

FIGURE 2 Estimated and Observed Fractions of Commodity Flows for Zone Choice Model of DC1, DC2, and DC3

TABLE 1 Summary of Input, Output, and Parameters to be Calibrated

Model		Input	Output	Parameters to be calibrated
Commodity Production and Consumption		Firm's characteristics such as: <ul style="list-style-type: none"> - Industry type. - Location. - Number of employees. - Floor area. - ect. 	1) Amount of commodity production by each firm. 2) Amount of commodity consumption by each firm.	$\alpha_1^k, \alpha_2^k, \dots, \alpha_n^k$ in Eq.(1) $\chi_1^k, \chi_2^k, \dots, \chi_m^k$ in Eq.(2)
Commodity Distribution	Distribution Channel Probability	Supply chain structure includes: <ul style="list-style-type: none"> - Industry type of a firm. - Location of a firm in a supply chain. 	Probability of a distribution channel being selected.	Calculate directly from the data
	Zone Choice Probability	1) Characteristics of zone such as <ul style="list-style-type: none"> - Total number of firms in a zone. - Total production commodities in a zone. - Population - Employment - Area 2) Spatial location of firms.	Probability of a zone being selected.	$\beta, \gamma, \rho, \sigma$ in Eq.(18) $\beta, \gamma, \rho, \sigma, \delta, \lambda$ in Eq.(19)
	Shipper Choice Probability	Characteristics of firm includes: <ul style="list-style-type: none"> - Industry type of a firm. - Total production commodities by a firm. - Total consumption commodities by a firm. 	Probability of a shipper being selected and finally get the commodity flows from shipper's firm to customer's firm.	Calculate directly from the data

TABLE 2. Percentage of Customer's Industry Type I_c Purchases Commodity Type k from Shipper's Industry Type I_s

Industry Type of Shipper (I_s)		Commodity Type (k)	Industry Type of Customer (I_c)			
			7	9	10	13
7	Other Manufacturer	6	67.3	46.1	31.1	23.2
9	Product Wholesaler	6	29.8	46.1	44.5	18.2
10	Retailer	6	2.9	7.9	24.4	43.8
13	Service & Government work	6	-	-	-	14.8
Total			100.0	100.0	100.0	100.0

TABLE3. Estimated Results for DC1, DC2, and DC3

Variables		DC1			DC2			DC3		
		MNL	SML-A	SML-AD	MNL	SML-A	SML-AD	MNL	SML-A	SML-AD
Zonal Attractiveness Variables										
NC (in 1,000's.)	parameter	0.298 ^{a)}	0.3751 ^{a)}	0.1829 ^{a)}	-	-	0.0687	0.0383	0.0578	0.0629
	t-value	(13.7)	(12.3)	(4.6)	-	-	(4.1)	(4.5)	(5.3)	(6.4)
AREA (km ²)	parameter	0.0039	0.0045	0.0032	0.0028	0.0036	-	-	-	-
	t-value	(13.3)	(13.0)	(7.4)	(10.7)	(8.6)	-	-	-	-
GEN (in 1,000,000 kg)	parameter	-	-	-	0.0026	0.0037	-	0.0108	0.0153	0.0075
	t-value	-	-	-	(15.2)	(10.3)	-	(26.9)	(15.4)	(6.0)
POP (in 10,000 persons.)	parameter	-	-	-	0.0052	0.0071	-	0.0030	0.0026	-
	t-value	-	-	-	(5.3)	(4.8)	-	(2.5)	(2.6)	-
EMP (in 10,000 persons.)	parameter	-	-	-	0.0073	0.0063	-	0.0588 ^{b)}	0.1609 ^{b)}	0.1569 ^{b)}
	t-value	-	-	-	(5.0)	(2.5)	-	(1.4)	(3.1)	(3.4)
Impedance Variables										
Distance (km)	parameter	-0.0623	-0.0736	-0.0631	-0.0429	-0.0582	-0.0500	-0.0249	-0.0337	-0.0300
	t-value	(-16.1)	(-14.9)	(-10.9)	(-19.7)	(-13.7)	(-12.3)	(-16.7)	(-14.5)	(-15.4)
Correlation Variables										
γ	parameter	-	-	-	-	2 ^{c)}	2 ^{c)}	-	2 ^{d)}	2 ^{d)}
	t-value	-	(69.0)	(64.0)	-	(-1.8)	(-1.5)	-	(-1.6)	(-1.6)
ρ	parameter	-	1.0168	0.9983	-	-1.3978	-4.8709	-	-11.673	-11.6057
	t-value	-	(69.0)	(64.0)	-	(-1.8)	(-1.5)	-	(-1.6)	(-1.6)
σ	parameter	-	-0.1177	-0.1121	-	1.5061	-2.3196	-	2.8734	2.5064
	t-value	-	(-2.0)	(-2.2)	-	(3.9)	(-4.4)	-	(4.0)	(3.8)
δ	parameter	-	-	0.4437	-	-	1.2169	-	-	1.5341
	t-value	-	-	(19.5)	-	-	(15.9)	-	-	(8.2)
λ	parameter	-	-	1.4823	-	-	1.5861	-	-	1.3001
	t-value	-	-	(11.2)	-	-	(6.1)	-	-	(3.9)
Number of observations		389	389	389	500	500	500	667	667	667
Log-likelihood at zero		-1537	-1537	-1537	-1841.3	-1841.3	-1841.3	-2635.5	-2635.5	-2635.5
Log-likelihood at convergence		-807.9	-773.9	-459	-1333.1	-1269	-1073.3	-2118.5	-2024.9	-1918.7
Adjusted likelihood ratio		0.474	0.496	0.701	0.276	0.311	0.400	0.196	0.232	0.272
AIC test		4.174	4.01	2.401	5.348	5.1	4.636	6.37	6.096	5.78

Note: ^{a)} NC 's unit is 1000's/km. ^{b)} EMP 's unit is 10,000 persons/km² ^{c),d)} The parameter is constrained to 2 for identification purposes