INDIVIDUAL-LEVEL STORE VISIT ANALYSIS USING A SPATIAL SEGMENTATION MODEL

Sotaro Katsumata*

Using store purchase record (panel) data, we propose a model to analyze individual store preference by incorporating store location and the residential area of the customer. The proposed model has the following three characteristics. First, since the model estimates parameters for each customer, it is possible to observe customer heterogeneity. The model incorporates not only individual heterogeneity, but also the time trend. Second, the model enables trade area analysis through the use of prior structure. By assuming a hierarchical structure for the model, it is possible to estimate the store visit probabilities of each geo-demographic segment. Third, by incorporating a spatial lag model into the prior structure of geographic parameters, it is possible to complement missing data to supply information from neighboring regions. Therefore, the cost of market survey research is likely to reduce with the application of the proposed spatial lag model.

1. Introduction

Trade area analysis plays an important role in marketing strategy planning for retail stores. Reilly (1929) proposed the law of retail gravitation, which analyses the visiting behavior of consumers in certain retail areas. This shows that researchers have long recognized the significance of trade area analysis in retail. Since the operations of one retail store strongly affect the performance of other neighboring stores, store managers have to deal in a competitive environment. Accordingly, they must analyze the trade area of their store and plan a competitive strategy to attract customers to it. Moreover, managers operating multiple retail stores must analyze the degree of cannibalization and competition among their own stores in order to implement a holistic corporate strategy.

In recent years, the market size of electronic commerce (EC) has increased rapidly with the wide proliferation of high-speed internet. However, even though the EC market is growing, it still falls behind the market size of real (face-to-face) trade. According to the Ministry of Economy, Trade and Industry in Japan (MIETI), retail EC sales make up approximately 2.5% of total retail sales in the country (MIETI, 2010). Such small numbers are not confined to Japan alone. Consider the United States, which is a bigger country than Japan, and which experienced a faster internet penetration. The EC sales rate in the United States is approximately 4.1% of total retail sales in the country (U.S. Census Bureau, 2010). Thus, even today, many transactions are conducted at the real store, and therefore, it is important to analyze

Key Words and Phrases: Trade Area Analysis, Spatial Lag Model, Hierarchical Modeling, Markov Chain Monte Carlo Method

* Faculty of Economics, Nagasaki University 4-2-1 Katafuchi, Nagasaki City, Nagasaki, Japan, 850-0003

Mail Address: katsumat@nagasaki-u.ac.jp
trade areas for real stores.

While trade area analysis continues to play an important role in retail, despite the advances in information technology, the utilization of store purchase records poses a challenge. In trade area analysis, customers’ addresses constitute important information. It is likely that the purchase frequencies of customers living at a distance from the store are less than those of customers residing near it. Some researchers try to collect addresses of customers visiting the store in order to examine the store’s trade area (e.g., Applebaum, 1964; Applebaum and Cohen, 1961). Nowadays, firms that offer a loyalty program also collect customers’ age and gender in addition to their addresses. These details allow firms to analyze the association between purchase behavior and customer address, and extract other valuable information. Purchase records, for example, can help identify customers who prefer stores located at a considerable distance from their residential areas, and offer customized one-to-one marketing offers to these customers. Thus, this data enables us to analyze both trade area (geographic segmentation) and customer heterogeneity.

Recently, some studies have analyzed trip data collected by the GPS (Global Positioning Systems). There is some potential for businesses to apply these technologies to marketing activities and social policies (Sekimoto et al., 2011). In fact, some firms try to utilize geographical information to formulate customer strategy (Nikkei MJ, 2013; 2014). In other cases, railway companies have introduced smart card ticketing systems that enable them to trace commuter transit and purchase information (Takamatsu et al., 2013). These recent cases imply that firms can obtain valuable implications not only from geographical distance data but also from individual behavioral data.

In this research, we construct a spatial lag model for retail trade area analysis. The model uses the purchase records of a firm operating more than one store and analyzes individual store visit behavior. Using the individual store visit record, we analyze the trade area in more detail and discuss possible management strategies for the retail organization in that particular area. We also discuss possible proactive applications of other findings obtained from the model.

2. Previous Research on Geographical Segmentation

In this section, we provide an overview of previous research in trade area analysis. We also discuss papers that studied the spatial lag model used to handle geographic information.

2.1 Spatial Lag Models

The spatial lag model (also called the spatial autoregressive model) was developed to examine spatial inter-correlation. The model is applied to various fields besides marketing, such as geographical or agricultural economics. Its basic properties and estimation methods are detailed in Anselin (1988, 2002).

We can define the spatial lag model as the following regression equation:
\[ y = \rho Wy + e, \quad e \sim N(0, \omega I) \]  

(2.1.1)

where \( y \) is a \( D \)-dimensional vector observation variable. Each element is the variable observed at each point, such as sales or the number of customers. In the model, there is a parameter \( \rho \) and an explanatory variable \( W \). \( W \) is a \( D \times D \) distance matrix, where all diagonal elements are 0. We define the row sum of the matrix \( W \) as 1, ; therefore, \( \sum_{c=1}^{d} w_{dc} = 1 \). \( \rho \) is a scholar parameter and its magnitude implies the impact from neighboring areas. \( \omega \) is a variance scholar parameter.

Then, by defining \( Q = (I - \rho W)^{-1} \), we can transform expression (2.1.1) as follows:

\[ y \sim N(0, \omega Q'Q) \]  

(2.1.2)

where \( Q' \) is transpose of \( Q \). When we use information from neighboring areas, we can construct a better model with fewer parameters by incorporating the spatial lag term. The model works well, especially for issues involving geographical factors. For example, Smith and LeSage (2004) examined the voting behavior of each state in the United States presidential election using a spatial lag model. Brueckner (2003) also adopted the model to analyze the interaction between local government policies.

This model has been applied to the marketing field not just for geographical segmentation but also for examining social networks, such as consumer communications (e.g., Iacobucci, 1996). Yang and Allenby (2003) proposed a social spatial lag model defining \( W \) as interdependent preference. Other relevant studies have been summarized in Bradlow et al. (2005).

Since this research also deals with spatial information, we construct a model incorporating a spatial lag. The next section provides a detailed explanation of the model.

### 2.2 Research in Trade Area Analysis

As mentioned above, Reilly (1929) proposed the retail gravity model in his pioneering work on trade area analysis. The retail gravity model is a theoretical model used to identify the borders of the trade area between cities. In its simplified form, the model estimates the sales volume of a customer by assuming two market centers (towns, shopping malls, etc.), which will divide the market between themselves. Customers will shop only at one center or the other (any crossover from one market is exactly offset by crossovers from the other). The division of the market can then be expressed as the product of the ratio of the sizes of the centers and the ratio of their distances from the boundary. Converse (1949) expanded this model to propose the concept of “dominant trade area,” whereby a city has the strongest retail gravity within its dominant area. Based on this study, Huff (1962, 1964) proposed a model to estimate store visit purchase probability, using which, we can obtain a probability contour curve map. Nakanishi and Cooper (1974) used a general formulation of Huff’s model to derive the MCI (multiplicative competitive interaction) model, which has been employed by many empirical studies (e.g., Black et al., 1985).
These models are used to determine the trading area using the population of the destination area, and distances from the residence to the destination area as input information. However, some studies have focused on investigating trade areas in a more direct way, such as by asking customers their area of residence. For example, Applebaum (1964) collected the residential addresses of store visitors, and defined two trade areas: primary and secondary. The primary and secondary trade areas were found to account for 60 to 70 percent, and 15 to 20 percent of total store sales, respectively. Berry (1967) also classified trade areas in a similar fashion.

More recently, some studies have conducted a geographical analysis of retail store locations. For example, Thomadsen (2007) analyzed the role of location and price competition in the fast food retail industry. Ter Hofstede et al. (2002) observed the factors affecting the international expansion of retail stores and analyzed consumer behaviors using data from European countries. Since these papers express consumer behaviors or responses and the factors affecting their outcomes as quantitative models, we can apply these proposed models to more general purposes. As pointed out by Nakanishi (1986) and Wedel and Kamakura (2000), we recognize that trade area analysis is a research subfield of market segmentation, which, among other things, focuses on geographic traits.

Recent research also throws light on other pertinent issues, which we address in this study. First, we deal with the demographic segment as a commonality of individual behavior. Marketing analysis has been based on the assumption of consumer heterogeneity, as evidenced by one-to-one marketing (Peppers and Rogers, 1993). Also, the hierarchical Bayes model has allowed practitioners to develop more realistic models of buyer behavior and decision making. Commonly used in marketing today, the model assumes prior distributions for individual parameters (e.g., Rossi, McCulloch, and Allenby, 1996; Rossi et al., 2005). The hierarchical Bayes model enables us to apply psychographic and demographic attributes to a prior distribution of a segment of consumers. Since these prior distributions reflect the commonality of consumers, we can deal with the market segment as a prior structure. For marketers, analyzing factors influencing individual behavior is especially desirable in constructing empirical models using scanner panel data. In addition, they have to consider incorporating time variation along with individual heterogeneity (Rossi and Allenby, 2003).

Second, we consider the particularity of the geographic factor. Geographic information is neither a linear (or log-linear) variable (such as age), nor a categorical variable (such as gender). Generally, neighboring geographical areas have similar characteristics unlike those that are distant from each other. Thus, we need a spatial lag model to formulate this assumption.

This research attempts to address the two above-mentioned issues.

3. Model

We construct a model to apply scanner panel data for a retail firm operating more than two stores. Consumer identification details (IDs) are common among stores and
3.1 Purchase Frequency

In many studies, the number (frequency) of purchases is assumed to follow a Poisson distribution (e.g., Schmittlein et al., 1987; Gupta and Morrison, 1991; Abe, 2009). This is particularly feasible for non-periodic purchased goods or aggregate goods. Frequency of purchase, an element of RFM analysis, is one of the most popular metrics used in database marketing (Blattberg et al. 2008). RFM analysis has three elements: Recency (R), Frequency (F), and Monetary value (M). In this study, we focus on the frequency of consumer purchase behavior. We will discuss the other elements of RFM analysis in section 6.

As previous studies suggest, frequency analysis is one of the most important issues in marketing. However, it is difficult to estimate parameters appropriately when the frequency of purchase visits is very low. Furthermore, the number of purchases may be affected both by consumer heterogeneities and seasonal factors. Thus, we need to incorporate these two factors into the model. Accordingly, we construct a model assuming that the marginal of each observation follows the Poisson distribution.

For store $j = 1, \ldots, J$, $y_{itj}$ is the number of purchases of customer $i = 1, \ldots, N$ at period $t = 1, \ldots, T$. Since the observation $y_{itj}$ may be 0, $y_{itj}$ is a natural number including 0. $y_{itj}$ follows a Poisson distribution with parameter $\exp(\lambda_{ij} + \kappa_{jt})$:

$$ y_{itj} \sim \text{Poi}(\exp(\lambda_{ij} + \kappa_{jt})) \quad (3.1.1) $$

where the range of parameters $\lambda_{ij}$ and $\kappa_{jt}$ are $(-\infty, \infty)$. Since the parameter is the product of exponential numbers, it satisfies the condition required for the Poisson distribution.

Since the Poisson distribution has a reproducing property, we obtain the following distributions from the marginal of each observation.

$$ \begin{cases} 
\sum_{t=1}^{T} y_{itj} \sim \text{Poi}(\exp(\lambda_{ij} \sum_{i=1}^{N} \exp(\kappa_{jt}))) \\
\sum_{i=1}^{N} y_{itj} \sim \text{Poi}(\exp(\kappa_{jt} \sum_{i=1}^{N} \exp(\lambda_{ij}))) 
\end{cases} \quad (3.1.2) $$

This means that the marginal distribution of the model also follows the Poisson distribution. Since there are sufficient marginal purchase, we can estimate the parameters from the sums of these observations, although we have to set constraints on parameters $\lambda_{ij}$ and $\kappa_{jt}$. An explanation for the same is presented in the Appendix.

3.2 Customer Heterogeneity and Spatial Lag

In this section, we explain the prior structure of the consumer heterogeneity parameter $\lambda_{ij}$. First, we assume that the store visit tendency for each consumer...
\[ \lambda_i = (\lambda_{i1}, \cdots, \lambda_{iJ})' \] can be explained as the following \( J \)-dimensional multivariate regression:

\[ \lambda_i = B x_i + \Delta r_i + \varepsilon_i, \varepsilon_i \sim \mathcal{N}_J(0, \Sigma) \] (3.2.1)

where \( x_i \) is a \( K \)-dimensional demographic variable vector, and \( B \) is a \( J \times K \) parameter matrix. \( \Delta r_i \) denotes the residential area effect of each consumer. Let the number of residential areas be \( D \), \( \Delta \) be a \( J \times D \) parameter matrix, and \( r_i \), a \( D \)-dimensional index vector. Since all consumers live in any one area, any one element of \( r_i \) is 1, and the remaining, 0. That is, if consumer \( i \) lives in area \( d \), \( r_{id} = 1 \), and \( r_{ic} = 0, c = 1, \cdots, d-1, d+1, \cdots, D \). \( \Sigma \) is a covariance matrix. Since the matrix parameter \( \Sigma \) reflects the competitive structure among stores, we can examine the interrelationships of stores.

We assume the prior structure of area factor \( \delta_d \) as the following spatial lag model, where \( \delta_d \) is \( d \)-th column of \( \Delta \) and a \( D \)-dimensional vector.

\[ \delta_j = \rho_j W (\delta_i - V_j \tau_j) + V_j \tau_j + \xi_j, \xi_j \sim \mathcal{N}_D(0, \omega_j I_D) \] (3.2.2)

where \( W_j \) is a \( D \times D \) distance matrix with the sum of its rows as 1. \( \rho_j \) is a scholar parameter that reflects the impact of the neighboring area. \( V_j \) is a \( D \times L \) matrix explanatory variable, such as store specific attributes. \( \tau_j \) is a \( L \)-dimensional parameter vector.

In the ordinary spatial lag model, \( \rho W \delta_j \) is only one element explaining \( \delta_j \). However, Anselin (2002) proposed a model to incorporate another explanatory variable. In this research, we further expand Anselin’s (2002) model. Let \( Q = (I - \rho_j W)^{-1}. \) Then, we can transform equation (3.2.2) as follows:

\[ \delta_j \sim \mathcal{N}_D(V_j \tau_j, \omega_j Q'Q) \] (3.2.3)

In this model, \( \delta_j \) follows a multivariate normal distribution with mean \( V_j \tau_j \), and variance \( \omega_j Q'Q \). Parameter \( \delta_j \) is affected by the neighboring area through the variance.

### 3.3 Time Series Term

For time factor \( \kappa_t = (\kappa_{t1}, \cdots, \kappa_{tJ})' \), we assume the following equation:

\[ \kappa_t = \Phi \kappa_{t-1} + \eta_t, \eta_t \sim \mathcal{N}_J(0, \Theta) \] (3.3.1)

where \( \Phi \) is a \( J \times J \) inertia parameter from a previous period. In this research, we fix it as a diagonal matrix. To simplify the discussion, we assume equation (3.3.1) to be a first-order autoregressive structure. However, we can assume a more complex structure by applying a time series model, such as that proposed by Dekimpe and Hansens (1999, 2000), Bronnenberg et al. (2000), Njis et al. (2001), and Dekimpe et al. (2008).

### 3.4 Model Summary

The proposed model is as follows:
\[
\begin{align*}
  y_{itj} & \sim \text{Poi}(\exp(\lambda_{ij} + \kappa_{jt})) \\
  \lambda_i & = Bx_i + \Delta r_i + \varepsilon_i, \varepsilon_i \sim N_J(0, \Sigma) \\
  \delta_j & = \rho_j W(\delta_i - V_j \tau_j) + V_j \tau_j + \xi_j, \xi_j \sim N_D(0, \omega_j I_D) \\
  \kappa_t & = \Phi \kappa_{t-1} + \eta_t, \eta_t \sim N_J(0, \Theta)
\end{align*}
\]

(3.4.1)

In our model, we have to estimate the following parameters: \( \lambda_{ij}, \kappa_{jt}, B, \Delta, \Sigma, \rho_j, \tau_j, \omega_j, \Phi, \) and \( \Theta \). We estimate them using the MCMC (Markov Chain Monte-Carlo) method (e.g. Gamerman, 1998; Gelman et al. 2003). A detailed explanation of prior and posterior distributions is described in the Appendix.

4. Empirical Analysis

4.1 Data Overview

For the empirical analysis, we use purchase history data of a firm operating three department stores (A, B, and C) in central Tokyo\(^1\). Since each customer has a unique ID, we can observe each customer’s purchase date and store. We also have details of age, sex, and residential area (municipal) of each customer. Therefore, we can roughly calculate the distance from their residence to each store. The data period covers 51 weeks.

We consider customers living within 30 kilometers of the nearest store as our sample. There are 107 municipals within the area, accounting for approximately 86% of total sales of these three stores. We can also cover the secondary trade area by referring to Applebaum (1964). The objective municipals are shaded dark grey in Figure 1\(^2\). The numbers denote the index of each municipal. The symbol “♦” indicates the locations of the three stores. It takes a customer between 30 minutes and an hour to travel to the nearest and farthest store, respectively, using public transportation.

4.2 Dependent and Independent Variables

In the empirical analysis, we define \( y_{itj} \) as the number of purchases of customer \( i (i = 1, \cdots, N) \), during the \( t \)-th week \( (t = 1, \cdots, T) \), at store \( j (j = 1, \cdots, J) \). The number of purchases is defined by the number of transactions. If a customer purchases more than one product at one visit to the shop, we count it as one purchase. If a customer makes purchases at two shops during one visit, we count it as two purchases.

For demographic variable \( x_i \), we use the customer’s age (logarithm) and gender (female = 1). Residential information is included in \( r_i \).

We set coordinates for each municipal as a city (or ward, town, or village) office. The \((c, d)\)-element of matrix \( W \) reflects the distance between each municipal. We set an inverse relationship for the distance. Therefore, if distance \((c, d)\) is small, \( W_{cd} \) is

\(^1\) However, we do not have detailed information regarding aspects such as floor space and target customers for the department chain and each store.

\(^2\) We used municipal boundary data provided by ESRI Japan, Inc. to draw the map. The boundary data is the copyright of ESRI Japan, Inc. (http://www.esrij.com/
large. Note that we confine the distance between a pair of municipals to 10 kilometers; if the distance between a pair of municipals exceeds 10 kilometers, we set $W_{cd}$ as 0 for those municipals. We set the sum of the rows of matrix $W$ to 1. That is, if the distance between municipals $c$ and $d$ is $\text{dist}(c, d)$, $W_{cd}$ would be as follows:

$$W_{cd} = \frac{d_{cd}}{\sum_{k=1}^{p} d_{kd}}$$

(4.2.1)

$$d_{cd} = \begin{cases} 
\frac{1}{\text{dist}(c, d)}, & \text{if } \text{dist}(c, d) \leq 10 \text{ km} \\
0, & \text{else}
\end{cases}$$

(4.2.2)

In this research, we define the explanatory variable on the mean of the spatial term as shown in equation (3.2.3). For variable $V_{dj}$, we use the intercept and inverse distance from municipal $d$ to store $j$, that is, $1/\text{dist}(c, j)$.

There are some other variables that define the distance, such as contiguous neighbors, ranked distance, and nearest (or n-nearest) neighbors (LeSage and Pace, 2008). In this study, we apply inverse distance, because distance can express the detailed geographic position of each municipality better than data on contiguous or nearest neighbors.
4.3 Calibration/Validation Sample

We divide the dataset into a calibration sample and a validation sample. We estimate the parameters of the proposed and comparison models from the calibration sample. Additionally, by using each estimated parameter, we predict the times of purchase using the validation sample and compare the predictive accuracy of the models. We set the maximum number of customers in one municipality to between 30 and 35. Then, we obtain the following values: overall number of calibration customers, $N = 3375$; number of stores, $J = 3$; time length, $T = 51$; and number of areas, $D = 107$. There are 2,803 female customers (83%) and 572 male customers (17%), with an average age of 27.7 years. As a summary of statistics, Figure 2 shows a color-coded map of the summary of objective variables ($y_{itj}$). The spectrum $(R, G, B) = (\bar{y}_{dA}/\bar{y}_d, \bar{y}_{dA}/\bar{y}_d, \bar{y}_{dA}/\bar{y}_d)$, where $\bar{y}_d = \sum_{j=1}^{J} \bar{y}_{dj}$. $y_{dj}$ is the average number of purchases per week by consumers who live in area $d$ and shop at store $j$.

Next, to compare the predictive accuracy of each model, we set the following dataset for validation. The validation sample will enable us to infer the purchase behavior of new customers. While we know where potential customers live within the validation area, as well as their sex and age, we do not know their prior purchase record. In
this paper, we use information from macro trends $\kappa_t, t = 1, \cdots, T$. However, as we are unable to directly observe $\lambda_i$ for $N^*$ new customers, we have to estimate their behavior using $Bx_i$ and $\Delta r_i$. Therefore, we devise a new customer set by selecting 409 customers separately from the calibration set. Then, $N^* = 409$. Let $y^*_t$ be an $N^* \times T$ matrix of the observed purchase history of new customers at store $j(j = 1, 2, 3)$. To test the predictive accuracy, we examine the difference between the predictive value and $y^*$.

We will first examine the in-sample/out-of-sample fit of following four models by using marginal likelihood and mean squared error (MSE). Thereafter, we will discuss the results. Since the expectation of the Poisson distribution is the same as its parameter, we will use the parameter as the predictive value and compare the four models in terms of these values. Suppose that the predictive value of customer $i(i = 1, \cdots, N^*)$ at time $t(t = 1, \cdots, T)$ at store $j$ is $\hat{y}_{itj}$, and the MSE is $(J \times T \times N^*)^{-1} \sum_{j=1}^{J} \sum_{t=1}^{T} \sum_{i=1}^{N^*} (\hat{y}_{itj} - y^*_{itj})^2$.

4.4 Comparison Models

In this research, we define the following proposed model (also referred to as the full model) and three other comparison models.

4.4.1 $\kappa$-model: time variety model

This model assumes the time variation of purchase frequency. However, it does not incorporate individual heterogeneity $\lambda_i$. In this model, we define times of purchase as $y_{itj} \sim \text{Poi}(\exp(\kappa_{tj}))$ and $\kappa_t = \Phi \kappa_{t-1} + \eta_t$.

To predict the purchase frequency of a customer in validation samples, we apply the parameter $\kappa_{tj}$. Therefore, for customer $i(i \in N^*)$, the expected purchase amount in $t$-th week is $\hat{y}_{itj} = \exp(\kappa_{tj})$. To obtain this predictive value, we use the posterior mean value of the parameter. The other models explained later also use the posterior mean value.

4.4.2 $\lambda$-model: customer heterogeneity model

This model assumes individual heterogeneity for purchase frequency. However, it does not incorporate time variety. We define purchase frequency as $y_{itj} \sim \text{Poi}(\exp(\lambda_{ij}))$. In addition, we set $\lambda_i = \tilde{B}\tilde{x}_i + \varepsilon_i$, for prior construct, where $\tilde{x}_i$ is a variable that adds the intercept term to $x_i$ and $\tilde{B}$ is a corresponding parameter. However, while this equation contains demographic variables, such as sex and age, it does not contain residential information.

To predict the purchase frequency of a customer in validation samples, we apply parameter $\tilde{B}$ and customer characteristics variable $\tilde{x}_i$. Therefore, for customer $i$ ($i \in N^*$), the expected purchase amount in $t$-th week is $\hat{y}_{itj} = \exp(\tilde{B}\tilde{x}_i)$.

4.4.3 $\delta$-model: customer geo-demographic heterogeneity model

This model assumes customer heterogeneity and the residential factor, but does
not include time variety. It incorporates the residential term and spatial prior construct to explain the parameter \( \lambda_i \). We set purchase frequency as \( y_{itj} \sim Poi(\exp(\lambda_{ij})) \).

Additionally, for prior construction, we set \( \lambda_i = Bx_i + \Delta r_i + \varepsilon_i \).

As in the case of the \( \lambda \)-model, to predict customer behavior, we apply parameter \( \tilde{B} \) and customer characteristics variable \( \tilde{x}_i \). In addition, we use the customers’ address information. For customer \( i(i \in N^*) \), the expected purchase amount in \( t \)-th week is therefore \( \tilde{y}_{itj} = \exp(\tilde{B}\tilde{x}_i + \Delta r_i) \).

4.4.4 Full model: proposed model

This model assumes both customer heterogeneity and time variation of purchase frequency. The number of purchases follows \( y_{itj} \sim Poi(\exp(\lambda_{ij} + \kappa_{ij})) \) and we set prior construction as \( \lambda_i = Bx_i + \Delta r_i + \varepsilon_i \), and \( \kappa_i = \Phi\kappa_{i-1} + \eta_t \).

The predicted purchase amount of a new customer at the \( t \)-th week is \( \tilde{y}_{itj} = \exp(Bx_i + \Delta r_i + \kappa_{ij}) \).

5. Result

5.1 Model Selection

At first, we will examine the fitness of models. Although, there are several indicators to compare models, in this research, we use the Bayes Factor (BF) to compare in-sample fitness. We have to obtain the marginal likelihood to calculate the BF. To do this, we use a method proposed by Newton and Raftery (1994). However, to compare out-of-sample fitness, we use the MSE.

The center column of Table 1 shows the results of the log marginal likelihood and log BF obtained. Log marginal likelihood values are on the diagonal elements and log BF values are on the lower triangle elements. The number in bold font indicates the highest value among the four models. For example, on the (2, 1)-element of the table, we can compare the \( \lambda \)-model and the \( \kappa \)-model. In this cell, the \( \lambda \)-model and \( \kappa \)-model are the proposed model and comparison model, respectively. We can see the extent of improvement in the fitness of the \( \lambda \)-model compared to the \( \kappa \)-model.

From the center column of Table 1, we infer that the log marginal likelihood value of the full model (proposed model) is the highest. We can see that the \( \delta \)-model follows the next. The fitness of the \( \delta \)-model has improved compared to that of the \( \lambda \)-model (log BF = 66.6). In Jeffreys (1961), if the BF exceeds 100 (roughly, log BF exceeds 4.61), the evidence against the comparison model is “decisive.” However, in Kass and Raftery (1995), if log BF exceeds 5, the evidence against the comparison model is “very strong.” Based on these thresholds and the in-sample fit, we can successfully improve the model by incorporating the residential factor.

Next, we compare the models using the validation sample (out-of-sample fitness). The right column of Table 1 shows the MSE between the forecast and observed values of each model. The bold text indicates the best value among the four models. From Table 1, the MSE of the full model is the lowest among the comparison models from
Table 1: Result of Model Comparison

<table>
<thead>
<tr>
<th></th>
<th>In-Sample-Fit: Log Marginal Likelihood and Bayes Factors</th>
<th>Out-of-Sample Fit: MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>κ-model</td>
<td>λ-model</td>
</tr>
<tr>
<td>κ-model</td>
<td>$-122,531.40$</td>
<td>0.3739</td>
</tr>
<tr>
<td>λ-model</td>
<td>23,481.30</td>
<td>$-99,050.10$</td>
</tr>
<tr>
<td>δ-model</td>
<td>23,547.80</td>
<td>66.60</td>
</tr>
<tr>
<td>Full model</td>
<td>25,691.00</td>
<td>2,209.70</td>
</tr>
</tbody>
</table>

the validation sample, indicating that the forecasting ability of the full model is the highest among all models. Although the values have not improved dramatically, the full model marks the best performance for all validation sets.

Therefore, the proposed model explains customers’ store purchase behavior most accurately. In the following section, we examine the estimated parameters of the full model.

5.2 Parameters of the Full Model

Table 2 shows the estimated result of parameters of the full model. The numbers without brackets indicate the posterior mean, while those enclosed within brackets denote the Highest Posterior Density (HPD) interval (2.5%, 97.5%). Additionally, “*,” “**,” and “***” indicate that 0 lies outside the 90%, 95%, and 99% HPD region, respectively. These highest posterior density intervals are calculated using the method proposed by Chen et al. (2000).

From Table 2, we can see the basic characteristics of each store from parameter B. For example, male customers tend to visit store A and B more frequently than females. On the other hand, female customers visit store C more frequently. Customer age profiles are clearly visible as well. Stores A and B are more frequently visited by younger customers, while this tendency is not discernible in Store C. We find that each store is preferred by a different segment of customers. Therefore, we must also consider that the trading areas of each store may differ depending on the demographic segments. We discuss this possibility in the next section.

From parameter Σ, we can examine the rivalry among the three stores. In Table 2, all the non-diagonal elements are negative and 0 lies outside the 99% HPD region. This means that many customers tend to visit one preferred store and do not visit other stores at all. We will discuss this finding in detail in the following section.

The $\rho$ values of all the spatial lag terms $s$ are close to 1. We can thus say that the effect from neighboring municipals is substantially strong and that the store visit tendencies among neighboring regions are similar. Also, parameter $\tau$ allows us to explore the tendencies of customers to visit stores depending on the distance. The parameter of the inverse distance between the residential municipal and the store (Distance) show that Stores A and B are preferred by customers living nearby, while customer tendency to visit Store C is not affected by this distance. The time series (inertia) term $\Phi$ shows, that for all stores, effects from previous periods are substantially high.
of the municipal areas correspond to those seen in Figure 1. Customers are placed at depicting customers living in municipal areas numbered 10, 13, 76, and 102. The IDs tendency of customers who live in certain municipal areas. There are four triangles ters. At first, we estimate individual store visit behavior. Fig. 3 shows the store visit 5.3 Trade Area

Figure 3 shows the time series transition of parameter $\kappa$. The solid line denotes Store A; the dotted line, Store B; and the dashed line, Store C. We find that all stores share one common tendency; sales at each store increased at about the same times several times in a year. This increase in sales occurs due to bargain sales. This department chain has two bargain sales. The first sale is in spring (roughly from the 16th to the 19th week of the year), and the second sale is in autumn (in roughly the 40th week).

So far, we have verified that we can improve the model by incorporating the time series term using the log BF values, as discussed in a previous section. We also verified a dramatic improvement in log BF values in the full model, when compared to the $\delta$-model, which did not include the time series term. Lastly, we confirmed that there are large time variations in one year.

5.3 Trade Area

In this section, we analyze the trade area of each store using the estimated parameters. At first, we estimate individual store visit behavior. Fig. 3 shows the store visit tendency of customers who live in certain municipal areas. There are four triangles depicting customers living in municipal areas numbered 10, 13, 76, and 102. The IDs of the municipal areas correspond to those seen in Figure 1. Customers are placed at

<table>
<thead>
<tr>
<th>$B$</th>
<th>Sex</th>
<th>Age</th>
<th>Store A</th>
<th>Store B</th>
<th>Store C</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\tau$</td>
<td>Intercept</td>
<td>Distance</td>
<td>$\omega$</td>
<td>$\Phi$</td>
<td>$\Theta$</td>
</tr>
<tr>
<td>$\rho$</td>
<td>$\omega$</td>
<td>$\Phi$</td>
<td>$\Theta$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Parameters of the full model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Store A</th>
<th>Store B</th>
<th>Store C</th>
</tr>
</thead>
<tbody>
<tr>
<td>$B$ Sex</td>
<td>$-0.456^{***}$</td>
<td>$-0.450^{***}$</td>
<td>$0.477^{***}$</td>
</tr>
<tr>
<td></td>
<td>$(-0.642, -0.274)$</td>
<td>$(-0.671, -0.226)$</td>
<td>$(0.332, 0.632)$</td>
</tr>
<tr>
<td>$B$ Age</td>
<td>$-1.414^{***}$</td>
<td>$-1.133^{***}$</td>
<td>$-0.129$</td>
</tr>
<tr>
<td></td>
<td>$(-1.626, -1.196)$</td>
<td>$(-1.360, -0.901)$</td>
<td>$(-0.298, 0.041)$</td>
</tr>
<tr>
<td>$\Sigma$ Store A</td>
<td>$2.619$</td>
<td>$(2.373, 2.846)$</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>$(-0.422^{***})$</td>
<td>$(2.323)$</td>
<td>$(1.541)$</td>
</tr>
<tr>
<td></td>
<td>$(0.715, 3.509)$</td>
<td>$(1.427, 1.662)$</td>
<td>-</td>
</tr>
<tr>
<td>$\rho$</td>
<td>$0.976^{***}$</td>
<td>$0.974^{***}$</td>
<td>$0.978^{***}$</td>
</tr>
<tr>
<td></td>
<td>$(0.918, 1.000)$</td>
<td>$(0.915, 1.000)$</td>
<td>$(0.924, 1.000)$</td>
</tr>
<tr>
<td>$\tau$ Intercept</td>
<td>$0.203$</td>
<td>$-2.716$</td>
<td>$-0.314$</td>
</tr>
<tr>
<td>$\tau$ Distance</td>
<td>$1.519^{*}$</td>
<td>$2.004^{***}$</td>
<td>$0.682$</td>
</tr>
<tr>
<td></td>
<td>$(-0.152, 3.251)$</td>
<td>$(0.886, 3.204)$</td>
<td>$(-2.487, 3.769)$</td>
</tr>
<tr>
<td>$\omega$</td>
<td>$0.709$</td>
<td>$1.329$</td>
<td>$0.349$</td>
</tr>
<tr>
<td></td>
<td>$(0.495, 0.956)$</td>
<td>$(0.944, 1.776)$</td>
<td>$(0.203, 0.518)$</td>
</tr>
<tr>
<td>$\Phi$</td>
<td>$0.437^{***}$</td>
<td>$0.500^{***}$</td>
<td>$0.509^{***}$</td>
</tr>
<tr>
<td></td>
<td>$(0.144, 0.727)$</td>
<td>$(0.217, 0.778)$</td>
<td>$(0.224, 0.799)$</td>
</tr>
<tr>
<td>$\Theta$</td>
<td>$0.122$</td>
<td>$0.128$</td>
<td>$0.123$</td>
</tr>
<tr>
<td></td>
<td>$(0.074, 0.175)$</td>
<td>$(0.079, 0.182)$</td>
<td>$(0.077, 0.176)$</td>
</tr>
</tbody>
</table>
the coordinates $l_{1i}/L_i, l_{2i}/L_i, l_{ij}/L_i$, where $l_{ij} = \exp(\lambda_{ij})$ and $L_i = \sum_{j=1}^{J} l_{ij}$. According to Figure 4, customers placed near the corners of the triangle tend to purchase exclusively at a particular store. For each area, “○” and “△” denote female and male customers, respectively, with the size of the symbol corresponding to the customer’s age. We can find that there are some general propensities for store preference. In areas numbered 10 and 76, many customers prefer Store C. Customers living in area number 13 prefer Store B, while those in area number 102 prefer Store A. However, there are some customers who prefer other stores. Therefore, if managers simply divide the trade area into segments based on the municipal areas, it is likely that their marketing actions targeted at these customers will be ineffective. Furthermore, as a general propensity, since many customers are located close to the corners, we can say that customers tend to prefer any one store, even if it is not the nearest to them from among the three stores. Moreover, customers positioned at the sides of the triangles visit two stores. Those located inside the triangles frequent three stores, although there are a few such customers.

We also find some heterogeneity among customers. For example, there are some customers in area number 76 who prefer Store A rather than Store C. Some of them seldom purchase products at Store C. For these customers, it is effective to implement customized one-to-one marketing rather than area-segmented marketing. Although some previous research (e.g., Kopp et al., 1989) shows the interrelationships between the competitive structure and the consumer segment, this result implies that there is substantial heterogeneity among consumers. Therefore, we have to examine individual behavior carefully; otherwise, it is possible that we may overlook the actual competitive structure.

Next, we will examine the aggregate tendency of store preference and analyze the trade area of each store. Since we can obtain store purchase tendency of municipal area $d$ from $\delta_d$, we will calculate the trade area of each segment with demographic

---

Figure 3: Transition of $\kappa$

---

\[ \text{coordinates: } l_{1i}/L_i, l_{2i}/L_i, l_{ij}/L_i, \text{ where } l_{ij} = \exp(\lambda_{ij}) \text{ and } L_i = \sum_{j=1}^{J} l_{ij}. \]
Figure 4: Individual store preference for some municipal areas

Note) Customers are placed at the coordinates $l_{i1}/L_i, l_{i2}/L_i, l_{i2}/L_i$, where $l_{ij} = \exp(\lambda_{ij})$ and $L_i = \sum_{j=1}^{J} l_{ij}$. “○” and “♦” denote female and male customers, respectively, with the size of the symbol corresponding to the customer’s age.

variables. For the individual demographic variable $x$, let $l_{dj} = \exp(B_j x + \delta_{dj})$ be the purchase frequency of a consumer whose demographic attribute is $x$ and residence is area $d$ for store $j$. Figure 5 is a four-color coded map of the trading area. The spectrum $(R, G, B) = (l_{dA}/L_d, l_{dB}/L_d, l_{dC}/L_d)$, where $L_d = \sum_{j=1}^{J} l_{dj}$. Figure 5 presents four demographic segments. As a general propensity, the trade area of Store A is the region southwest to the store, while that of Store B lies to its northwest, and that of Store C covers the northern and eastern regions. In addition, Stores A and C compete in the southwest bay area. Moreover, we identify the store trade area for each demographic attribute, such as age and gender. For example, Figure 5 shows that older female customers prefer Store C.

When firms with access to individual purchase records try to estimate their trade area based on data collected from customers, it is desirable to identify customers who
exhibit store choice behavior distinct from that of the majority. Although, firms tailor their marketing communications for customers, acting on this finding can give firms a competitive advantage. Since their competitors do not have access to the individual purchase records of a particular store’s customers, they cannot customize communications to fit individual preference. In addition, in order to analyze the aggregate level trade area, firms have to incorporate other segmentation traits, such as demographic attributes, into area-specific parameters.

6. Discussion

6.1 Complementing Missing Information

Since we construct a model incorporating a spatial lag term, we discuss the advantage of the spatial lag structure. The prior structure of $\delta$ in the proposed model is defined in equations (3.2.2) and (3.2.3). By eliminating the spatial lag from the prior
structure, we can obtain following equation:

$$\delta_j = V_j \tau_j + \xi_j, \xi_j \sim N_D(0, \omega_j I) \Leftrightarrow \delta_j \sim N_D(V_j \tau_j, \omega_j I)$$  \hspace{1cm} (6.1.1)

If information of a municipal area is missing, we can instead use its neighbor’s information as a replacement for $\delta$. Therefore, we can say that the spatial prior is insusceptible to missing data. Then, we examine to what extent the spatial prior can act as an accurate substitute for the missing information compared to its non-spatial counterpart defined in equation (6.1.1). The number of objective areas in this research is 107. From these, we randomly exclude some areas and estimate $\Delta$. Then, we examine the degree of misfit from the parameter $\Delta$ estimated by using the complete dataset. We exclude customers living in the area from this dataset and supply information from neighboring areas.

Figure 6 shows the accuracy of the complementation of missing information. We change the rate of the missing area from 10% to 90% and estimate parameters for each case. At the same time, we also estimate $\Delta$ from equation (6.1) for the non-spatial prior. To examine the complementation accuracy, we use the correlation coefficients of the complete and missing datasets for parameter $\Delta$. We obtain the correlation for each store and calculate the average among the three stores. Let parameter $\Delta$ obtained from the dataset of missing rate $p$ be $\Delta^{(p)}$. Figure 6 shows the calculated correlation variables from $\frac{1}{3} \sum_{j=1}^{3} \text{cor}(\delta_j^{(0)}, \delta_j^{(0.1)})$ to $\frac{4}{3} \sum_{j=1}^{3} \text{cor}(\delta_j^{(0)}, \delta_j^{(0.9)})$, where $\delta_j^{(p)}$ is the $j$-th column of $\Delta^{(p)}$, which indicates that the parameter of store $j$, and $\text{cor}(v_1, v_2)$ is a correlation coefficient between $v_1$ and $v_2$. A higher correlation value implies that the model can fill missing data more accurately.

We find that the complementation accuracy of the spatial lag prior is higher than the non-spatial prior for all cases. Furthermore, differences in accuracy between the two models become greater as the missing rate increases. Thus, the spatial prior is robust against missing data.

The cost of new market survey research is likely to reduce with the application of the proposed spatial lag model. There is no need to conduct a survey of all municipal
areas within a target region. The spatial lag model also enables us to complement the information of municipal areas outside the scope of a market survey. Figure 7 shows a rough image of the model’s characteristics. In the spatial prior model, managers can complement information from the neighboring area. However, in the non-spatial prior model, they complement information based on common prior parameters and do not refer to the neighboring area. As a result, the spatial prior model provides a more accurate estimate of area characteristics than the non-spatial prior model. In such an area, shown in Figure 7, managers have to conduct a market survey among all 9 areas if they use the non-spatial prior model to obtain each area’s characteristics. On the other hand, if they use the spatial prior model, since the model can complement information, they can obtain enough reliable area characteristics without conducting market research in all 9 areas. Therefore, the spatial prior model helps firms to reduce market survey costs.

6.2 Implications for Geographic Segmentation

These empirical results imply that the proposed model has several advantages over the previous geographic segmentation models. For example, from the result shown in Figure 4, we can find some customers who do not follow the general tendency of customers. In other words, a simple geographic segmentation model that estimates customers’ store choice behaviors only from resident data cannot provide an accurate estimation. For example, the Huff model of trade area gravity (Huff, 1962) estimates the attractiveness of an area from the distance to and the population of that area. This model cannot assume customer heterogeneity.

Undoubtedly, we owe our models to advances in information technology. However,
firms have to utilize their own databases to maintain customer relationships and thus achieve a competitive advantage. When firms want to maximize the lifetime value of each individual customer, they have to determine the heterogeneity of customers. The proposed model estimates consumers’ store visit behavior using not only geographic information but also their store choice records and time-trends. Therefore, the model offers a feasible way to assess customer characteristics.

Further, from the results shown in Figure 5, we can infer that prior information allows firms to extract inferences regarding area characteristics. The general tendency of each area helps them acquire new customers. We also find that the trade area of each store does not correspond to the straight-line distance. It seems that the trade areas follow public transportation such as the railway. Therefore, it is feasible to estimate store choice tendencies from customer purchase records rather than geographical information such as distance. Our model can be applied to geographic segmentation issues.

6.3 Expansion of the Model

As Abe (2009) and Blattberg et al. (2008) observe in their studies, customer value is obtained from frequency of purchase and the monetary value of the purchases made. We can easily calculate the sales amount by multiplying the frequency and monetary value. Although we have proposed a model to estimate frequency of purchase in this paper, firms have to estimate monetary value as well. Therefore, in this section, we discuss how to expand the model to estimate monetary value.

One approach to estimate monetary value is to apply the Type I Tobit model (Amemiya 1984). We can observe the monetary value of the purchases of customer \(i\) at \(t\)-th week if the customer goes shopping at least one time. Otherwise, we cannot observe the value. Therefore, the Tobit model is feasible. In addition, since some researchers have proposed a way to estimate parameters using the MCMC method, we can easily obtain the parameters (Albert and Chib, 1993; Koop, 2003). Further, we can assume the interaction between frequency of purchase and monetary value.

For example, let \(z_{itj}\) be an observation variable, and \(z_{itj}^*\) be the latent variable (Koop 2003). If we assume that \(z_{itj}^* = \lambda_{ij}^* + \kappa_i^* + \epsilon_{itj}\) and \((\lambda_i, \lambda_i^*)' = Bx_i + \Delta r_i + \epsilon_i, \epsilon_i \sim N(0, \Sigma)\), where \(\lambda_i^* = (\lambda_{i1}, \cdots, \lambda_{iJ})'\), the part of the matrix parameter \(\Sigma\) become an interaction term between \(\lambda_i\) and \(\lambda_i^*\).

Extending on previous research, LeSage and Polasek (2008) propose a model to explain the interregional flows using the spatial model. This is also applicable to some empirical issues in database marketing.

7. Conclusion

In this research, we proposed a model to analyze individual store preference by incorporating store location and the residential area of the customer using purchase record data. Our research made the following three contributions.
First, we constructed a model to estimate customer heterogeneity. The proposed model assumed scanner panel data, such as customer purchase record, as an input data, and separately estimated individual-specific and time variation factors. We could improve the fitness of the model by incorporating not only individual heterogeneity, but also the time trend.

Second, we applied the prior structure of the model to trade area analysis. Since the proposed model has a hierarchical structure, we incorporated demographic factors or residential information. Using the structure, we obtained the general store visit propensity for each market segment, such as age, gender, and residential area. We also obtained an understanding of the store trade area from this result. It is useful to obtain this information in order to acquire new customers. We found that the trade area is not completely determined by the distance of the customer’s residence from the store. Additionally, we found that the range of the trade area differs depending on demographic variables such as age or gender of the customers. Therefore, firms must tailor their range of marketing communications based on the demographic attributes of their target consumers.

Third, the proposed model is highly robust against missing information. In fact, the spatial lag model can complement missing data to supply information from neighboring regions. The proposed model can therefore be applied to market survey research.

In this study, we do not have detailed information on the objective department chain and on each individual store in that chain. Therefore, our discussion is limited. Firms can obtain further implications to apply the model to their own customer database. In addition, it is desirable for firms to analyze rival customer store choice tendencies in each area if this information is available.

Furthermore, since our spatial lag model can incorporate explanatory variables into the mean value, the model may be further refined to allow the use of effective variables. We therefore propose the following for future research. First, we aim to improve the time series term to a more effective form. As mentioned in section 3.3, many studies have proposed useful empirical models along these lines. We can make further refinements to the model based on their findings. Second, we have to consider an application to estimate the trade area of a new store. We found that although some stores held an edge over the competition, it was not apparent how a new store might compete against existing stores, and how it would affect their trade areas. To this end, we may expand the method proposed by Applebaum (1964), and incorporate experiential information into the competitive structure parameter $\Sigma$ and store specific parameters $\rho$, $\tau$.

Acknowledgements

The author would like to thank Prof. Makoto Abe (the University of Tokyo) for advice and feedback, Joint Association Study Group of Management Science for providing the data.
Appendix

We detail the estimating procedure below. Although, we have estimated additional models beyond the proposed model, and conducted comparisons among them, we prove that the additional models are actually simplified versions of the proposed model. Thereafter, we explain the proposed or full model only.

A.1 Likelihood Function and Prior Distributions

Recall the model:

\[
\begin{align*}
  y_{itj} & \sim \text{Poi}(\exp(\lambda_{ij} + \kappa_{jt})) \\
  \lambda_i & = Bx_i + \Delta r_i + \varepsilon_i, \varepsilon_i \sim N_J(0, \Sigma) \\
  \delta_j & = \rho_j W(\delta_j - V_j \tau_j) + V_j \tau_j + \xi_j, \xi_j \sim N_D(0, \omega_j I_D) \\
  \kappa_t & = \Phi \kappa_{t-1} + \eta_t, \eta_t \sim N_J(0, \Theta)
\end{align*}
\]

For prior distribution, we assume that

\[
B \sim N_{I \times K}(B_0, \Sigma, S_{B0}), \Sigma \sim W(s_{\Sigma 0}, S_{\Sigma 0}), \tau_j \sim N_M(\tau_0, S_{\tau 0}), \omega_j \sim G(s_{\omega 0}/2, S_{\omega 0}/2), \text{diag}(\Phi) = \phi \sim N_J(\phi_0, S_{\phi}), \text{and } \theta_j \sim G(s_{\theta 0}/2, S_{\theta 0}/2).
\]

For \(\rho_j\), we assume that \(\rho_j \sim U(c_{\min} - 1, c_{\max} - 1)\), following Smith and LeSage (2004), where \(c_{\min}\) and \(c_{\max}\) are the minimum and maximum eigenvalue of the distance matrix \(W\), respectively. Note that \(c_{\min} < 0\) and \(c_{\max} > 0\). From these prior distributions, we can define the full conditional posterior distribution as follows:

\[
\pi(\vartheta|D) = f(Y|\Lambda) f(\Lambda|B, \Delta, \Sigma) f(\Delta|\rho, \tau, \omega) \pi(\rho) \pi(\tau) \pi(\omega) \pi(B|\Sigma) \pi(\Sigma) f(K|\Phi, \Theta) \pi(\Phi) \pi(\Theta)
\]

where

\[
f(Y|\Lambda) = \prod_{i=1}^N \prod_{t=1}^T \prod_{j=1}^J \pi(y_{itj}|\lambda_{ij}, \kappa_{jt})
\]

\[
f(\Lambda|B, \Delta, \Sigma) = \prod_{i=1}^N \pi(\lambda_i|B, \Delta, \Sigma)
\]

\[
f(\Delta|\rho, \tau, \omega) = \prod_{j=1}^J \pi(\delta_j|\rho, \tau, \omega)
\]

\[
f(K|\Phi, \Theta) = \prod_{t=1}^T \pi(\kappa_t|\kappa_{t-1}, \Phi, \Theta)
\]

In this research, we set \(B_0 = O_{I \times K}, S_{B0} = 100I_J, s_{\Sigma 0} = K + D, S_{\Sigma 0} = 100I_J, \tau_0 = 0, S_{\tau 0} = 100I_M, s_{\omega 0} = K, S_{\omega 0} = 100I_J, s_{\theta 0} = J, \text{and } S_{\theta 0} = J\). For the initial value of the time series term, we assume that \(\kappa_{0j} = 0\). Therefore, \(\kappa_{1j} = \eta_t, \eta_t \sim N_J(0, \Theta)\). To solve the identification problem, we introduce the restriction \(\kappa_{Tj} = 0\).
A.2 Posterior Distributions

A.2.1 Posterior distribution of $\lambda_{ij}, i = 1, \cdots, N, j = 1, \cdots, J$

We obtain the following distributions:

$$
\pi(\lambda_{ij} | \cdot) \propto \prod_{t=1}^{T} \pi(y_{itj} | \lambda_{ij}, \kappa_{jt}) \pi(\lambda_{i} | Bx_i, \Delta, \Sigma) \\
\times \prod_{t=1}^{T} \left[ \exp(\lambda_{ij} + \kappa_{jt})^{k_{ij}} \times \exp(-\exp(\lambda_{ij} + \kappa_{jt})) \right] \\
\times \exp \left\{ -\frac{1}{2}(\lambda_i - Bx_i - \Delta r_i)'\Sigma^{-1}(\lambda_i - Bx_i - \Delta r_i) \right\}
$$

(A.2.1)

Since this is not a well-known distribution, we have to obtain samples using the Metropolis–Hastings (M–H) method. In this research, we obtain samples using the random-walk M–H method. A candidate sample $\lambda^N$ is generated from a normal distribution with mean is previous sample $\lambda^O$. Therefore, $\lambda^N \sim N(\lambda^O, s^2)$, where $s$ is an adjustment parameter. We set $s^2 = 0.1$. The acceptance ratio of the candidate sample $a$ is defined as follows:

$$
a = \min \left\{ 1, \frac{\pi(\lambda^N | \cdot)}{\pi(\lambda^O | \cdot)} \right\}
$$

(A.2.2)

A.2.2 Posterior distribution of $\kappa_{tj}, t = 1, \cdots, T - 1, j = 1, \cdots, J$

The conditional posterior distribution is as follows:

$$
\pi(\kappa_{tj} | \cdot) \propto \prod_{i=1}^{N} \pi(y_{itj} | \lambda_{ij}, \kappa_{tj}) \pi(\kappa_{tj} | \kappa_{t-1,j}, \Phi, \Theta) \pi(\kappa_{t+1,j} | \kappa_{tj}, \Phi, \Theta)
$$

(A.2.3)

Similar to $\lambda_{ij}$, we obtain samples using the random walk M–H method. The sample generating procedure is also the same as that for $\lambda_{ij}$.

A.2.3 Posterior distribution of $B$

Let $\Lambda = (\lambda_1, \cdots, \lambda_N)'$, $X = (x_1, \cdots, x_N)'$, and $R = (r_1, \cdots, r_N)'$. The regression equation of $\lambda_i$ is rewritten as the following matrix equation:

$$
\Lambda = XB' + R\Delta' + E, E \sim N_{N \times J}(0, I_N, \Sigma)
$$

(A.2.4)

From equation (A.10), we can obtain the conditional posterior as a matrix normal distribution

$$
B | \cdot \sim N_{J \times K}(B_1, \Sigma, S_1)
$$

(A.2.5)

where $S_1 = (X'X + S_0^{-1})^{-1}$, and $B_1 = ((\Lambda - R\Delta')X + S_0^{-1}B_0)S_1$. Refer to Dawid (1981) or Rowe (2002) for further details of the matrix normal distribution.
A.2.4 Posterior distribution of $\delta_j$

We obtain the posterior sample of $\Delta$ for each column $\delta_j$. First, we subtract the $j$-th column from expression (A.10) to obtain

$$\lambda_j = X\beta_j + R\delta_j + e_j, e_j \sim N_N(\mu_j, s_jI_N) \quad (A.2.6)$$

where $\mu_j = (\lambda_j - X\beta_j - R\delta_j)^{-1}\Sigma_{-,j}^{-1}\Sigma_{-,j}$ and $s_j = \Sigma_{jj} - \Sigma_{j,-j}^{-1}\Sigma_{-,j}$. From this equation, we obtain the following posterior distribution:

$$\delta_j | \cdot \sim ND(d_1, S_{\delta}) \quad (A.2.7)$$

where

$$S_{\delta} = (s_j^{-1}R'R + (\omega_jQ'Q)^{-1})^{-1}, d_1 = S_{\delta}(s_j^{-1}R'f_j - (\omega_jQ'Q)^{-1}V_j\tau_j), Q = (I - \rho_jW)^{-1}, \text{ and } f_j = \lambda_j - (X\beta_j + \mu_j).$$

A.2.5 Posterior distribution of $\Sigma$

The posterior sample of $\Sigma$ can be obtained from the following Wishart distribution:

$$\Sigma^{-1} | \cdot \sim W(s_{\Sigma}, S_{\Sigma}^{-1}) \quad (A.2.8)$$

where $s_{\Sigma} = s_{\Sigma}0 + N + K$ and $S_{\Sigma} = S_{\Sigma}0 + (\Lambda - XB' - R\Delta')'(\Lambda - XB' - R\Delta') + (B - B_0)\Sigma_{B0}^{-1}(B - B_0)'$.

A.2.6 Posterior distribution of $\rho$

The conditional posterior distribution of $\rho$ is as follows:

$$\pi(\rho_j | \cdot) \propto \pi(\rho_j) \times \pi(\delta_j | \rho_j, \tau_j, \omega_j) \propto (c_{\min}^{-1} - c_{\max}^{-1}) \times \exp \left\{-\frac{1}{2}(\delta_j - V_j\tau_j)'(\omega_jQ'Q)^{-1}(\delta_j - V_j\tau_j)\right\} \quad (A.2.9)$$

where $Q = (I - \rho_jW)^{-1}$. Since the posterior distribution is not well known, we obtain samples using the random walk M–H method. Although we obtain samples using this method, the range of samples must be $\rho_j \in (c_{\min}^{-1}, c_{\max}^{-1})$ (Sun et al., 1999; Smith and LeSage, 2004). Then, we generate a candidate sample $\rho^N$ from the truncated normal distribution with mean $\rho^O$. Therefore, $\rho^N \sim TN(c_{\min}^{-1}, c_{\max}^{-1}) (\rho^O, s^2)$, where $s^2$ is an adjustment parameter, and we set $s^2 = 0.1$. The density of the candidate distribution is defined as follows:

$$q(\rho^N | \rho^O) = \frac{\phi \left( \frac{\rho^N - \rho^O}{s} \right)}{s \left[ \Phi \left( \frac{c_{\max}^{-1} - \rho^O}{s} \right) - \Phi \left( \frac{c_{\min}^{-1} - \rho^O}{s} \right) \right]} \quad (A.2.10)$$

where $\Phi(\cdot)$ is a cumulative distribution function of the standard normal distribution, and $\phi(\cdot)$ is a density function. Using this equation, acceptance ratio $a$ is as follows:

$$a = \min \left\{1, \frac{\pi(\rho^N | \cdot)q(\rho^N | \rho^O)}{\pi(\rho^O | \cdot)q(\rho^O | \rho^N)} \right\} \quad (A.2.11)$$
A.2.7 Posterior distribution of \( \tau_j \) and \( \omega_j \)

The posterior sample of \( \tau_j \) and \( \omega_j \) can be obtained from the following distributions:

\[
\tau_j | \cdot \sim N(\tau_1, S_{\tau 1}) \]  \hspace{1cm} (A.2.12)

where \( S_{\tau 1} = (S^{-1}_{\tau 0} + V'_j(\omega_j Q'Q)^{-1}V_j)^{-1}, \) \( \tau_1 = S_{\tau 1}(S^{-1}_{\tau 0}\tau_0 + V'_j(\omega_j Q'Q)^{-1}\delta_j), \) and \( Q = (I - \rho_j W)^{-1}. \)

\[
\omega_j | \cdot \sim G(s_\omega 1/2, S_\omega 1/2) \]  \hspace{1cm} (A.2.13)

where \( s_\omega = s_\omega 0 + D, S_\omega = S_\omega 0 + (\delta_j - V_j\tau_j)'Q'Q(\delta_j - V_j\tau_j). \)

A.2.8 Posterior distribution of \( \phi_j \) and \( \theta_j \)

Since \( \Phi \) and \( \Theta \) are both diagonal matrices, the posterior sample is obtained for each element.

\[
\phi_j | \cdot \sim N(\phi_1, S_{\phi 1}) \]  \hspace{1cm} (A.2.14)

where \( S_{\phi 1} = (S^{-1}_{\phi 0} + \theta^{-1}_j\sum_{t=1}^T\kappa^2_{t-1,j})^{-1} \) and \( \phi_1 = S_{\phi 1}^{-1}(S^{-1}_{\phi 0}\phi_0 + \theta^{-1}_j\sum_{t=1}^T\kappa_{t-1,j}\kappa_{t,j}). \)

\[
\theta_j | \cdot \sim G(s_\theta 1/2, S_\theta 1/2) \]  \hspace{1cm} (A.2.15)

where \( s_\theta = s_\theta 0 + T \) and \( S_\theta = S_\theta 0 + \sum_{t=1}^T(\kappa_{t,j} - \phi_j\kappa_{t-1,j})^2. \)

Strictly speaking, our model should be interpreted as a state space model. This time series term corresponds to the transition equation, while the first term of expression (A.1.1) corresponds to the measurement equation. However, to estimate parameters in a state space model, we have to consider that the measurement equation is not the same as in a normal linear regression model. One of the issues that we hope to examine in future is that of finding an appropriate procedure to estimate this type of state space model.

A.2.9 Simulation Settings

In this study, we estimate a proposed model and three comparison models. For all four models, we obtain 15,000 samples after 10,000 iterations. We find that all the parameters converge to stable distributions.

REFERENCES


Huff, D. H. (1962). Determination of Intra-Urban Retail Trade Areas, *Real Estate Research Program, Graduate School of Business Administration, Division of Research, University of California, Los Angeles*.


(Received January 29 2014, Revised June 16 2014)