MMRC
DISCUSSION PAPER SERIES

No. 409

A brand purchase model of consumer goods incorporating the information search and learning process

Sotaro Katsumata
Nagasaki University

Makoto Abe
The University of Tokyo

May 2012

東京大学ものづくり経営研究センター
Manufacturing Management Research Center (MMRC)

A brand purchase model of consumer goods incorporating the information search and learning process

Sotaro Katsumata
(Nagasaki University)
Makoto Abe
(The University of Tokyo)

Abstract

In this research, we use scanner panel data to construct a stochastic brand choice model of consumer goods in which consumers repeatedly choose a brand from many alternatives. We thus examine consumers' repeat purchase behavior from the perspective of information processing theory. In particular, we explicitly incorporate the concept of internal search, external search, and learning, which have been proposed in behavioral studies, into the presented brand choice model. Previous research on brand choice has suggested the existence of choice subsets, such as an "awareness set" or a "consideration set," in the minds of consumers when they make a purchase decision. These subsets cannot be observed directly from purchase data, however, because their identification requires either direct questioning or inference through behavioral modeling. Instead, in this study, we introduce the concept of an "experiential set" as a means for consumers to process information and decide on brand choice. Crucially, the experiential set is observable from purchase records. Because choice subsets are constructed from observable data, this concept helps build brand choice models that incorporate more elaborate information searches and learning processes by consumers. This, in turn, results in a high predictive validity of the model.

Keywords: Information Search, Discrete Choice Model, Consideration Set, Markov Chain Monte Carlo Method

1. Introduction

In order to meet the needs of various customers, firms tailor their products to different market segments. Consumers are thus able to choose the most suitable product from the brands available in the market. As the number of brands available in the market increases, theoretically, consumers should be able to obtain higher utility. In practice, however, consumers do not evaluate all the brands that exist in the market nor do they choose brands through rational decision making. For example, even though the shampoo market comprises over 100 brands with each brand providing different benefits, few consumers can evaluate all these brands when deciding what to buy. Thus, in reality, consumers do not exercise rational choice behavior as assumed by microeconomic theory (Simon, 1947). In mature markets, most firms and brands face this circumstance.

Simon (1947, 1997) introduces the concept of "bounded rationality" in which many alternatives and problems exist in the real world. He also proposes a decision process whereby consumers do not evaluate all alternatives but review only a subset of them in order to choose the most preferred option. The imperfection in human cognition is a serious concern in Marketing. Some marketing models assume that consumers allocate cognitive resources, such as time and effort, differentially across brands when forming their attitude. For example, the Howard-Sheth model assumes consumers' inner process of brand comprehension and attitude formation through environmental stimuli and learning (Howard and Sheth, 1969). Petty and Cacioppo (1986) propose the Elaboration Likelihood Model (ELM) which posits that consumer's information process differs according to his or her level of product knowledge and involvement. These models support the notion that consumers do not evaluate all brands equally.

Although the theories of bounded rationality and selective information processing provide ample marketing implications, it is difficult to incorporate them into an empirical analysis of consumer purchase behavior because of data limitations. For example, scanner panel data only tell us what brand was purchased by whom and when. We cannot investigate which brands were evaluated before the customer made his or her actual purchase. From the previous argument, it is clear that consumers consider only a subset of brand alternatives. However, unless survey research such as direct questioning is carried out, it is difficult for firms to know this "bounded" subset. If we were to incorporate this inner process into a choice model, we are faced by the issue of inferring consumers' brand subsets from observed purchase data. In this paper, we thus construct a brand purchase model that incorporates a brand subset formation process from purchase data alone.

2. The Theory of Brand Choice

2.1. Bounded Rationality in Brand Choice and Brand Subsets

Many of the brand choice models used in marketing are founded on the framework of stochastic utility maximization. Thus, they assume that consumers have utilities for all brands. When the number of available brands is large, however, some consumers may not be aware of or interested in certain brands. For these brands, therefore, utility does not exist.

While the concept of bounded rationality (Simon, 1947, 1997) was pioneering in proposing the idea of a brand subset, many conceptual models of subset formation in the field of marketing have been proposed based on the cognitive aspects of consumers. For example, Howard and Sheth (1969) define the "evoked set" as a subset of available brands, and only these brands are evaluated by consumers. Narayana and Markin (1975) classify brands into inept and inert types. Some studies have introduced the type of subset, such as the "choice set" defined by Hauser and Shugan (1989) or the "consideration set" proposed by Wright and Barbour (1977) and Roberts (1989). In addition, others have applied these subsets in a multi-stage decision process (e.g., Lapersonne, Laurent and Le Goff, 1995; Brisoux and Cheron, 1990).

A major difficulty in operationalizing these models is the fact that we cannot obtain information on brand subsets except by directly asking consumers which brands they included. Shocker et al. (1991) and Robert and Lattin (1991) propose approaches that allow researchers to infer a brand subset from behavioral (purchase) data. Andrews and Srinivasan (1995) extend these models by estimating brand subsets stochastically, and this concept has since been followed by other studies, such as Chiang, Chib, and Narashimhan (1999), Gilbride and Allenby (2004), and Nielop et al. (2010). For practical use, however, these models have serious constraints. As the number of brands n increases, the possible number of brand subsets increases exponentially as $2^n - 1$. In many product categories, n is much greater than 10, implying that the number of subsets is unmanageable.

In summary, we cannot observe intermediate brand subsets directly from behavioral data. Although some models attempt to estimate these subsets stochastically, their application is limited to cases that have only a few brands. In section 2.2, we examine the formation of brand subsets from the perspectives of information searching and brand screening by consumers.

2.2. Steps in Brand Choice from the Perspective of Information Processing

Information processing models are theoretically based on the S-O-R model, such as the Nicosia model (Nicosia, 1966), Howard–Sheth model (Howard and Sheth, 1969), and EKB model (Engel, Kollat and Blackwell, 1968). These models examine a consumer's inner purchase decision process. Further development leads to information processing models of motivated consumers, such as that pioneered by Bettman (1979) and followed by other studies (e.g., Mitchell, 1981; Howard, Shay and Green, 1988). These models assume that consumers have some sort of goal or need. In addition, Blackwell, Miniard, and Engel (2006) apply information processing theory to improve the EKB model and thus propose the consumer decision process model.

The Bettman and consumer decision process models share two common structures. First, when consumers have a motive to solve a problem, they evaluate the available alternatives using knowledge stored in their internal memories. They resort to searching outside information only when they are dissatisfied with the alternatives evaluated using their memories. Second, when consumers purchase a brand, they learn from its usage, and its experience feeds back to their long-term memories.

Following these structures, search and consumption processes can be divided into the following three steps (e.g., Hoyer and MacInnis, 2008; Mowen, 1995):

1. Internal search: first, consumers search their internal memories to solve the problem.

2. External search: if consumers cannot solve the problem through this internal search, they refer to outside information.
3. Learning: after consumption, this experiment is stored in consumers' internal memories, and on the next purchase occasion, an internal search is executed based on this updated memory.

These internal and external search mechanisms are similar to the concept of bounded rationality. An internal search is conducted within the knowledge of consumers that is "bounded" in comparison to all brands available in the market. Information processing models further assume that an external search should be carried out when consumers are unsatisfied with the results of the internal search. Howard and Sheth (1969) also assume that the feedback system in that purchase experience affects satisfaction and brand comprehension.

2.3. A Subset of Experience and Learning from Repeat Purchase Behavior

In this section, we examine how to incorporate the feedback system into a brand choice model. Because we cannot observe consumers' inner information processing decision from purchase data, we must somehow infer this search and learning.

Consider the case when consumer i purchases brand j on the t-th purchase occasion. If brand j was purchased previously, consumer i must have knowledge of it. Therefore, previously purchased brands are evaluated by an internal search. By contrast, if brand j has not been purchased before, this brand is presumably evaluated by an external search. Having purchased and consumed brand j, the brand is now stored in the memory. Thus, on the (t+1)th purchase occasion, brand j will become an element of the brand subset that is evaluated by the internal search.

From purchase data, we can obtain the set of brands stored in the memory of each consumer (i.e., those included within the internal search). This subset is not a set of favorable brands. We name this the "experiential set" and define it as follows. The experiential set is a subset of brands that is formed through repeat purchases and evaluated by an internal search. It consists of brands that have already been purchased and used by the consumer. It is conceptually different from the "consideration set" and the "choice set," which exist before choice and from which one brand is selected to be purchased. The "experiential set" is formed after choice as a result of learning and memory storage feedback. The brand purchased on the next occasion may not necessarily come from the experiential set.

By introducing this concept of the experiential set, we are able to incorporate the feedback system of consumer information processing explicitly into the brand choice model using observable data only. Fig. 1 shows the purchase process described above.

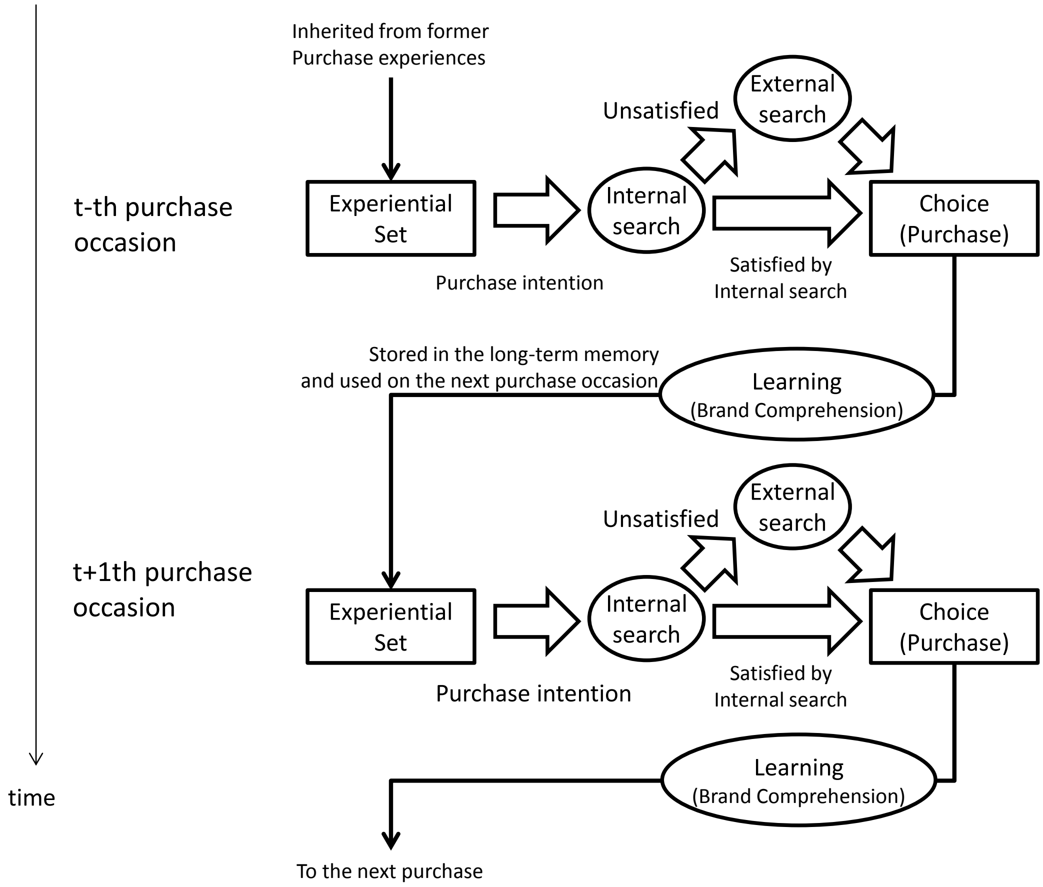

Fig. 1. The model of repetitive search and learning

One advantage of introducing the experiential set is that we can obtain the time-series formation of brand subsets explicitly from purchase data. The importance of examining dynamics in brand subsets was raised as early as the late 1970s (Farley and Ring, 1974).

3. The Experiential Set Purchase Model

In this section, we formulate the brand purchase model that incorporates the concept of the internal search, external search, and experiential set based on the search and learning model as the extension of ordinary brand purchase model discussed in section 2.

3.1. The Experiential Set

Consumers store information on specific brands in their long-term memories. This brand set—the experiential set—is constructed through past purchase behavior. In this section, we construct the purchase model that incorporates the experiential set.

First, let E_{it} be the experiential set of consumer i on the t-th purchase occasion. Because the experiential set changes over time, t is attached. The number of brands within the set is defined as $n(E_{it})$. In the following sentence, we develop the experiential set along with the relationship with the observable variables.

When consumer i opts to purchase a product on the t-th occasion, s/he first conducts an internal search. At this time, the consumer retrieves brand information from E_{it}, which was formed after the $(t-1)$th purchase occasion. If the consumer purchases the brand, which is a member of E_{it}, at t, we see that consumer i finds a satisfactory brand from the internal search and purchases this brand. In this case, the experiential set on the $(t+1)$-th occasion is the same as the set at t, that is, $E_{i,t+1} = E_{it}$. By contrast, if the consumer remains unsatisfied following the internal search, s/he would find an alternative through the external search. When consumer i purchases brand j, which is not in the set E_{it}, on the t-th purchase occasion $(j \notin E_{it})$, $E_{i,t+1}$ contains the element of E_{it} and brand j also becomes a member of the set. In this case, the experiential set expands following the external search.

Let us define the observable variables. First, when consumer i purchases brand j on the t-th purchase occasion, let the purchase output variable $y_{itj} = 1$. In this case, brand j is a member of the experiential set $(j \in E_{it})$ and brand k, which is a member of E_{it} and is not purchased, we define as $y_{itk} = 0$ $(k \neq j, k \in E_{it})$. The external search is observed when the purchased brand j is not a member of E_{it}. In other words, the external search is conducted when $j \notin E_{it}$; thus, let the external search variable $z_{it} = 1$. In this case, we observe the first purchase of brand j; however, because brand j is not a member of the set E_{it}, we do not use this occasion to estimate the purchase probability of brand j. As described in the definition of the experiential set E_{it}, this is the set of brands that have already been purchased and used. Therefore, the experiential set E_{it} expands not only at the point of purchase but also after the purchase.

Table 1 shows the constructs of the processes of the internal search, external search, and experiential set. This table depicts the process of a consumer who has the experiential set only containing brand a at the initial state (start observation, $t = 0$). The consumer expands his/her experiential set through the external search. In this table, the first purchase is denoted by (1), which means that the first purchase occasions cannot be used to estimate brand purchases. In this example, we include the case that the consumer purchases nothing or purchases more than two brands simultaneously.

Purchase occasion t	Brand purchase a	b	c	Experiential set E_{it}	External search Z_{it}
0				{a}	0
1	1			{a}	0
2	1	(1)		{a}	1
3	0	0		{a, b}	0
4	0	0		{a, b}	0
5	0	0	(1)	{a, b}	1
6	1	0	0	{a, b, c}	0
7	0	1	0	{a, b, c}	0
8	1	0	1	{a, b, c}	0

Table 1. The example of search behavior and the experiential set

3.2. Formulation of the Model

In this section, we assume the constructs in order to explain the observed variables y and z. Because these variables take $\{0, 1\}$ discrete values, we introduce the discrete choice model. In this research, we follow Albert and Chib (1993) in order to construct a model that has continuous latent variables. This way of formulation is an application of the data augmentation proposed by Tanner and Wong (1987) using the Markov Chain Monte Carlo (MCMC) method. We use this method to estimate parameters.

At first, let introduce the latent variable y^*_{itj}, which corresponds y_{itj}. These two variables have following condition:

$$y_{itj} = \begin{cases} 1 & \text{if} \quad y^*_{itj} > 0 \\ 0 & \text{if} \quad y^*_{itj} \leq 0 \end{cases} \tag{1}$$

The latent variable y^*_{itj} represents the utility of consumer i for brand j at t-th purchase occasion. For brand j, in order to be chosen by consumer i at t, latent variable y^*_{itj} must be exceed a certain threshold and we let the threshold be 0. In the t-th purchase occasion, when all of the latent variables of brands which are the member of the subset E_{it} below 0, no one brands are purchased at this occasion. On the other hand, when the latent variables of more than two brands exceed 0, all of these brands are purchased. This structure of model is called the multinomial model (e.g. Chib and Greenberg, 1998; Manchanda, Ansari and Gupta, 1999) as contrasted with the multinomial model (e.g. McCulloch and Rossi, 1994, 1996; Allenby and Rossi, 1999; McCulloch, Polson and Rossi, 2000). This model is distinguished what distribution are assumed on the error term (e.g. Train, 2003). If the distribution of the error term is the extreme value distribution, the model is called the "logit model", while the error term is normally distributed, the model called the "probit model". In this research, since we assume the normal distribution, the type of model is the "multivariate probit model".

As same as the observation of brand purchase y_{itj}, we also introduce the continuous latent variable z^*_{it}, which corresponds z_{it} as follows:

$$z_{it} = \begin{cases} 1 & \text{if} \quad z^*_{it} > 0 \\ 0 & \text{if} \quad z^*_{it} \leq 0 \end{cases} \tag{2}$$

Using above latent variables, we assume the construct to explain these brand choice and external search behavior. First of all, let x_{it} be the K dimensional explanatory variable such as the sales promotion reached consumer i at the t-th occasion. x_{it} includes the intercept. We assume that degree of the response to these variables are differ from each brand and customer, so we assume the coefficient as β_{ij}. The coefficient β_{ij} includes the degree of response of the intercept means the brand-dummy. From the above variables, we assume the following brand choice model:

$$y^*_{itj} = x'_{it}\beta_{ij} + \varepsilon_{it}, \qquad j \in E_{it}, t = 1, \dots, T_i \tag{3}$$

Additionally, in this research, we assume the correlation among each brand. Let J be the number of brands available in the market (number of whole brand set), we define Σ be the $J \times J$ parameter of correlation matrix. The correlation Σ holds full information of correlation structure of focal market, however, we assume that each consumer do not know full information of the market. For consumer i, at the t-th purchase occasion, the consumer refers the corresponding subset of Σ. We denote this submatrix as $\tilde{\Sigma}_{it}$ which consists of the correlation elements of brands which are member of the experiential set E_{it}. In the same manner, we let $\tilde{y}^*_{it}$ is a vector whose elements are corresponding brands which are member of E_{it} extracted from y^*_{it}. Also, $\tilde{B}_{it}$ is $n(E_{it}) \times K$ matrix which is extracted corresponding elements from $J \times K$ matrix $B_i = \{\beta_{i1}, \cdots, \beta_{iJ}\}'$. Using these variables, the brand purchase model is defined as following $n(E_{it})$ dimensional multivariate regression model:

$$\tilde{y}^*_{itj} = \tilde{B}_{it}x_{it} + \tilde{\varepsilon}_{it}, \quad \tilde{\varepsilon}_{it} \sim N(0, \tilde{\Sigma}_{it}), \quad t = 1, \dots, T_i \tag{4}$$

To construct the external search model, we define following regression equation using explanatory variable w_{it} and its coefficient parameter γ_i:

$$z^*_{it} = w'_{it}\gamma_i + \zeta_{it}, \quad \zeta_{it} \sim N(0, 1), \quad t = 1, \dots, T_i \tag{5}$$

In this research, we assume the hierarchical construct on parameter B_i and γ_i to discuss the relation between each parameter and demographic traits. This is one of reason to use MCMC method. Let v_i be a demographic variable vector, a matrix parameter Δ, and an individual parameter $\theta_i = \{\mathrm{vec}(B_i)', \gamma'_i\}'$. W, we assume the following relation:

$$\theta_i = \Delta v_i + \xi_i, \quad \xi_i \sim N(0, \Phi) \tag{6}$$

In the equation (6), we need the brand dummy and response parameter for all brands. However, in brand purchase equation (4), we estimate only for the brand which is a member of the experiential set E_{it}. Therefore, the parameters of brands which are not an element of the experiential set of the last occasion E_{i,T_i} are missing values. In MCMC method, since we are able to estimate these missing values stochastically, we use the algorism and estimate these values.

In the next section, we will show the explanatory variables in detail after overview the empirical data. Also, detailed description of prior distribution, posterior distribution, and other settings of the model is provided in the appendix.

4. Empirical Analysis

4.1. Data Overview

For the empirical analysis, we use the sales records of a drugstore chain provided by the *Joint Association Study Group of Management Science* and *Customer Communications*. The period of data covers two years from January 1 2008 to December 31 2009. The studied product category is shampoo, which is a low price commodity that is generally purchased repeatedly. Because of these properties, it is suitable to assess our proposed model.

4.2. Empirical Analysis Setting

4.2.1.Period of Analysis and Studied Brands

First, to apply the proposed model, we need a period before beginning the analysis in order to assess the initial state of the experiential set. Therefore, in this research we use the first 9 months of the 24-month study period in order to form the initial experiential set. The period of analysis is thus 15 months.

The initial experiential set contains the brands that were purchased in the first 9 months (this is called the formation period). Additionally, to assess the predictive accuracy of the model, we exclude the last purchase occasion of each customer from the analysis (i.e., the validation period). The remaining period of data is called the calibration period.

We choose the top 10 brands purchased within the whole study period as our study objects. Although the computation load of the proposed model does not dramatically increase as the number of studied brands increases, the reliability of the estimation results will decrease. Table 2 shows the 10 studied brands and their sales.

	brand name	sales amount
1	Lux	1406
2	Pantane	1226
3	TSUBAKI	734
4	merit	710
5	Essential	640
6	Dove	546
7	Super Mild	447
8	Soft in One	376
9	PB (Private Brand)	420
10	Mod's Hair	389

Table 2. List of Studied Brands

4.2.2. Definition of the Purchase Occasion and Studied Customers

The proposed model has a multivariate choice structure in which more than two brands can be chosen on the same purchase occasion and a consumer can choose nothing at all. We define the purchase occasion based on this structure. Thus, we set the day when the consumer purchases a shampoo product as the purchase occasion.

The purchase probability is the conditional probability of a category purchase occurring. Although we do not discuss this concept in detail, we are able to obtain the purchase probability on any day in order to estimate the category purchase probability and the brand purchase probability obtained by our proposed model. Some researchers have proposed a model that divides purchase probability into store visit, category purchase, and brand purchase (e.g., Chiang, 1991; Chintagunta, 1993; Chib, Seetharaman, and Strijnev, 2004; Van Heerde, and Neslin, 2008). We can extend the proposed model to apply these models.

The number of studied consumers is 400. These are randomly chosen from consumers who purchase over three times both in the formation period and in the calibration period.

4.2.3. Variable Definition

For the explanatory variable x_{it}, this vector includes the intercept, sale day dummy, weekend and holiday dummy, and the size of the experiential set. Because in this drugstore, the first and twentieth day of each month are sale days when all products are discounted, we define a sale day dummy variable. Furthermore, we use the size of the experiential set as one of the explanatory variables. For explanatory variable w_{it}, in this research, we use the same variables as x_{it}.

4.2.4. Comparison Model

We construct another model in tandem with the proposed model, namely the comparison model. For this, we use the logit model, which is estimated using the most likelihood method. The purchase probability of consumer i purchasing brand j on the t-th purchase occasion is defined as follows:

$$\Pr\left(y_{itj} = 1\right) = \frac{\exp(x'_{it}\beta_j)}{1 + \exp\left(x'_{it}\beta_j\right)} \tag{7}$$

where

$x_{it} = \{\text{Intercept}, \ [\text{Sale day}]_t, \ [\text{Weekend and Holiday}]_t, \ [\text{Gender}]_i, \ [\log(\text{Age})]_i \}$.

We obtain the parameter for each brand and use this for the prediction.

5. Results

5.1. Parameter Convergence and Prediction

First, we assess the convergence of the parameters of the proposed model. To discriminate the convergence, we apply the method proposed by Geweke (1992). We use the first 10% and the last 50% of the sample sequence to test the differences between both sample sequences. As a result, we confirm that all parameters are converged.

Using these parameter samples, we forecast the purchase behavior of each customer during the validation period. We obtain the purchase probability of the $\tau_i = (T_i + 1)$-th purchase and use this score for the presented predictions.

First, the purchase probability of brand j by consumer i on the τ_i-th purchase occasion is obtained from the following equation when brand j is a member of the experiential set E_{i,τ_i}, where the experiential set at $\tau_i = T_i + 1$ is available using purchase records until T_i and $\Phi(a)$ is the probability function of the standard normal distribution evaluated as

$$\hat{p}_{i\tau_i j} = \Pr\left(\hat{y}_{i\tau_i j} = 1\right) = \Phi(\hat{y}^*_{i\tau_i j}), \ \text{if} \ \ j \in E_{i\tau_i} \tag{8}$$

We have to consider the relationships between the focal brand and the other brands. To obtain $\hat{y}^*_{i\tau_i j}$, we generate a random sample from $\left(\hat{y}^*_{i\tau_i j}\right)^{(h)} \sim N\left(\tilde{B}_{i\tau_i} x_{i\tau_i}, \tilde{\Sigma}_{i\tau_i}\right)$ and take the mean of samples where $\tilde{B}_{i\tau_i}$ and $\tilde{\Sigma}_{i\tau_i}$ are the partial matrices of B_i and Σ, which contain elements of the corresponding brands in the experiential set $E_{i\tau_i}$.

The external search probability is obtained from the following equation:

$$\hat{q}_{i\tau_i} = \Pr\left(\hat{z}_{i\tau_i} = 1\right) = \Phi(\hat{z}^*_{i\tau_i}) \tag{9}$$

In the case of the purchase probability of brand j, when brand j is a member of the experiential set, we can calculate the probability straightforwardly. However, when brand j is not a member of the set, we cannot obtain the purchase probability in a precise sense because the proposed model only estimates the external search probability—it cannot estimate which brand will be chosen after the external search. However, in the real world, it is desirable to forecast the purchase behavior of brands outside the experiential set.

Therefore, in this research we apply and compare the following two forecasting methods. The first method sets the probability as 0 for the brands outside the experiential set in order to obey the theory faithfully and what the model describes. The second method obtains the probability of the brands outside the experiential set by multiplying the purchase probability estimated from the prior structure by the external search probability. Although this method is slightly different from the rigorous theory of the model, we are able to obtain the purchase tendency of all brands of some sort.

The purchase probability estimated from the prior structure is obtained from the demographic variable v_i and its prior parameter Δ. The predictive score is obtained from the following equation, where $\breve{B} = \Delta v_i$ and $\breve{y}^*_{itj} = \breve{\beta}_{ij} x_{i\tau_i j}$:

$$\hat{p}_{i\tau_i j} = \hat{q}_{i\tau_i} \times \Phi(\breve{y}^*_{i\tau_i j}) \tag{9}$$

We now compare the predictive accuracy of the two methods above with the comparison model (logit). To measure the predictive accuracy, we use the hit rate, ROC (receiver operating characteristic) curve, and ROC score. The hit rate indicates the matching rate between the prediction and the observation. We set the threshold as 0.5. The ROC score has a value between 0 and 1; as this value approaches 1, it implies that it has an improved predictive accuracy. A model has a good predictive accuracy when the score exceeds 0.5. Detailed descriptions are provided in Blattberg, Kim, and Neslin (2008).

Table 3 shows the forecasting result. In the table, the "number of buyers" means the number of consumers who purchase the focal brand during the validation period. In the external search, the value shows the number of consumers who purchase the brand outside the experiential set. The "non-purchase rate" means the rate of consumers who do not purchase the focal brand during the validation period. This value is used as the benchmark of the hit rate. As Table 3 shows, the non-purchase rates of the studied brands are higher than 0.5. Furthermore, if the forecasting method suggests that no consumers will purchase, the hit rate will be equivalent to the value. Therefore, we have to compare the hit rate of each method with the non-purchase rate. The method is said to have a predictive ability only if the hit rate exceeds the non-purchase rate.

In Table 3, “method 1” sets the purchase probability as 0 if the brand is not a member of the experiential set, whereas “method 2” estimates the purchase probability of brands that are outside the experiential set from prior information and the external search probability. The boldface variables indicate that the method marks the highest performance of the three (in the external search, of the two).

			Hit rate			ROC score		
	number of buyers	non-purchase rate	comparison model	method1	method2	comparison model	method1	method2
Lux	73	0.854	0.854	**0.952**	**0.952**	0.574	0.965	**0.977**
Pantane	76	0.848	0.848	**0.916**	**0.916**	0.563	0.911	**0.936**
TSUBAKI	49	0.902	0.902	**0.972**	**0.972**	0.627	0.973	**0.977**
merit	56	0.888	0.888	**0.944**	**0.944**	0.620	0.941	**0.966**
Essential	37	0.926	0.926	**0.970**	**0.970**	0.601	0.935	**0.947**
Dove	39	0.922	0.922	**0.968**	**0.968**	0.598	0.920	**0.922**
Super Mild	23	0.954	0.954	**0.970**	**0.970**	0.624	0.911	**0.949**
Soft in One	29	0.942	0.942	**0.984**	**0.984**	0.586	0.928	**0.952**
PB (Private Brand)	30	0.940	0.940	**0.974**	**0.974**	0.611	0.925	**0.959**
Mod's Hair	13	0.974	0.974	**0.992**	**0.992**	0.522	**0.999**	**0.999**
external search	39	0.922	0.922	0.922		0.592	**0.926**	

Table 3. Predictive accuracy

Note) the results of the external search of methods 1 and 2 are the same.

Table 3 shows that the comparison model has a predictive ability to some degree because the ROC scores of all brands exceed 0.5. However, the hit rates of all brands are the same as the non-purchase probability, which implies that the comparison model cannot discriminate between potential buyers and non-buyers. By contrast, the hit rates of both methods using the result of the proposed model of all brands exceed the non-purchase probability. Furthermore, the ROC scores of all brands exceed 0.5, and they are higher than the comparison model. In particular, the predictive performance of method 2 is higher than method 1 for most brands.

We show in Fig. 2 the ROC curves that were used to obtain the ROC scores for certain brands. We are able to check visually whether the method has a predictive ability by viewing the ROC curve. If the area under the ROC curve exceeds 0.5, or the ROC curve exceeds the 45-degree line (dotted line), we can say that the model (method) has a predictive ability. Compared with the comparison model, the predictive performances of both two methods of the proposed model are fairly high. In particular, method 2 outperforms method 1 for middle- to low-level consumers. This means that method 2 has a predictive ability for the purchase tendencies of consumers who do not contain the focal brands in their experiential sets.

In summary, the proposed model has a high predictive ability and thus the model is applicable for analyzing the sales data.

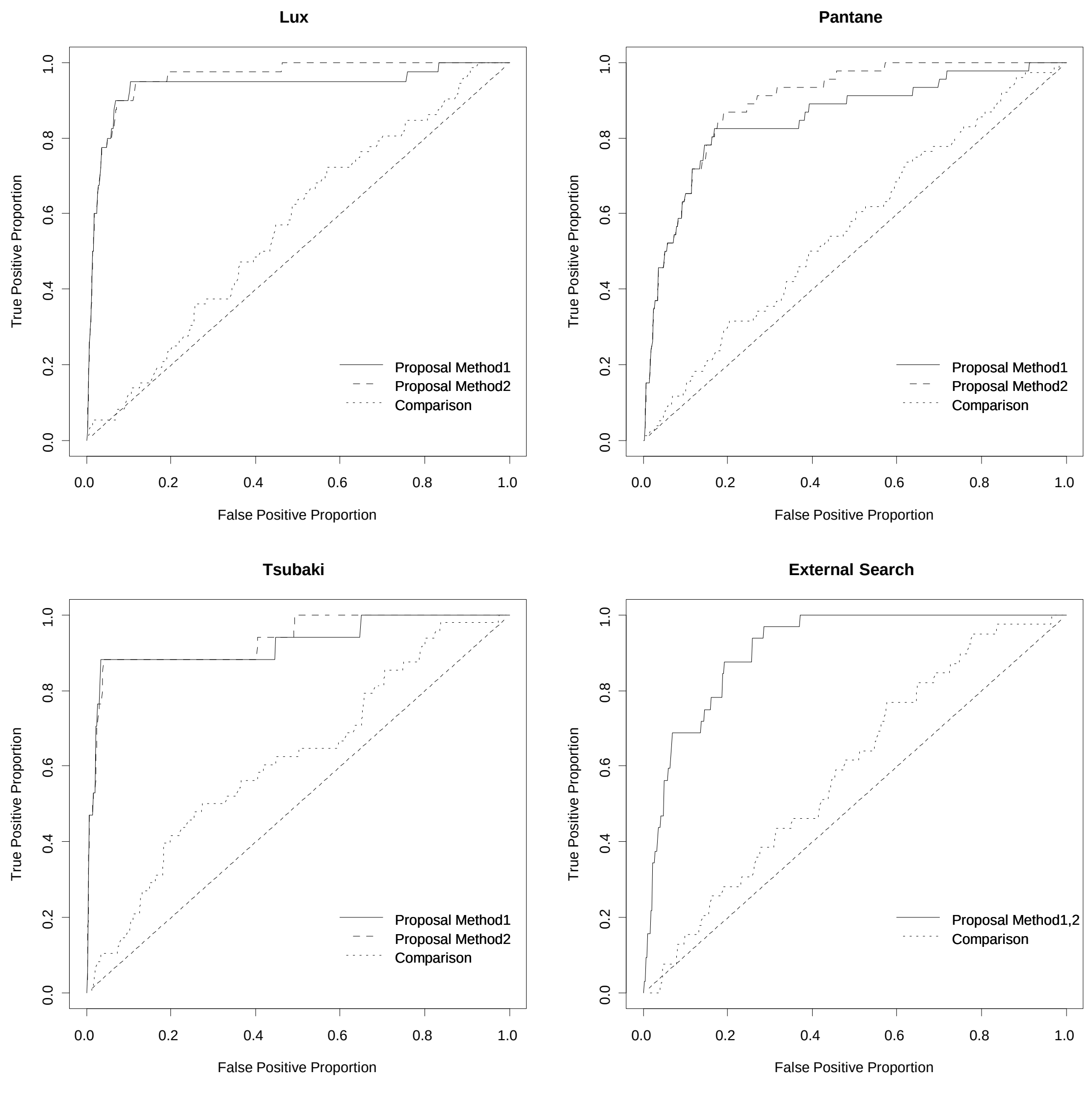

Fig. 2: ROC curves

5.2. Model Parameters

We next infer the general tendency of consumers in order to assess the parameters of each explanatory variable. Table 4 shows the selected estimation result of parameter Δ. In the table, "*," "**," and "***" indicate that 0 lies outside the 90%, 95%, and 99% highest posterior density intervals of the estimate. These highest posterior density intervals are calculated using the method proposed by Chen, Shao, and Ibrahim (2000).

We found some differences among each brand. First, from [Intercept - Gender], we see that Super Mild is favored by female customers, whereas Soft in One is favored by older male consumers. Furthermore, from [Sale day - Intercept], the purchase probability of Tsubaki increases on sale days. By comparing these characteristics of each brand, we can formulate a brand communication strategy. For example, Super Mild initiates the concept "Super Mild cheers for fathers who take a bath with their children.[1]" This concept is for male customers who have younger children.

		Intercept	Gender	Age
Tsubaki	Intercept	5.02	0.80	-1.06
	Sale day	7.54 **	-5.05 ***	-0.40
	Weekend and Holiday	3.50	-2.58	-0.11
	Size of the experiential set	-0.73	0.64	-0.29
Super Mild	Intercept	-2.09	-3.59 ***	1.43
	Sale day	0.67	0.40	-0.20
	Weekend and Holiday	3.76	0.27	-0.98
	Size of the experiential set	-0.95	1.73 ***	-0.31
Soft inOne	Intercept	-13.53 **	-3.30 *	4.43 ***
	Sale day	-2.65	-8.35 **	2.70
	Weekend and Holiday	-6.04	2.00	1.26
	Size of the experiential set	1.15	1.36 *	-0.78

Table 4: Estimated Δ

1 SHISEIDO Co., Ltd., http://www.shiseido.co.jp/sm/concept/index.html, retrieved on May 21, 2012.

Table 5 reports the posterior mean of Σ. As in Table 4, "*," "**," and "***" indicate that 0 lies outside the 90%, 95%, and 99% highest posterior density intervals of the estimate. We find that some elements are negatively significant. This means that when one brand is chosen, another brand tends not to be. In other words, these two brands are in a competitive relationship. However, the estimated Σ shows the brand interrelationships of the whole market, implying that no individual consumer has complete information about the market structure. This matrix is estimated for each consumer and for the assembled competitive relationships of the whole market.

	Lux	Pantane	TSUBAKI	Merit	Essential
Pantene	-0.066 *				
TSUBAKI	-0.035	-0.023			
Merit	-0.036	-0.033	-0.026		
Essential	-0.007	0.0033	-0.016	-0.008	
Dove	-0.02	-0.02	0.0017	-0.012	-0.04
Super Mild	-0.01	-0.01	-0.013	-0.043	-0.028
Soft in One	-0.032	-0.018	-0.073	-0.033	-0.03
PB (Private Brand)	-0.011	-0.007	-0.013	-0.029	0.0279
Mod's Hair	-0.007	-0.009	-0.024	-0.05	0.006

	Dove	Super Mild	Soft in One	PB
Pantene				
TSUBAKI				
Merit				
Essential				
Dove				
Super Mild	-0.009			
Soft in One	0.064	0.03		
PB (Private Brand)	-0.039	-0.027	-0.014	
Mod's Hair	0.0105	-0.001	-0.023	-0.013

Table 5: Estimated Σ

6. Discussion

6.1. Understanding the Parameters

In this research, we estimate the parameters of brands that are members of the experiential set. This means that the obtained estimates are the brand preferences of consumers who have experience of using the brand in question. If a brand is purchased by a consumer through an external search and it thus becomes a member of the experiential set, its preference value will be lower if the consumer does not purchase the brand thereafter. Therefore, parameters are directly affected by the use values of consumers aside from the expected value derived from sales promotions or advertisements. In this section, we discuss the application of the parameters B_i and Δ, which have above properties.

First, let us consider the case of the preference of brand j by consumer i who has not yet purchased brand j yet. If the preference value of brand j that is obtained from the hierarchical structure is high, is the consumer more likely to choose brand j over other brands during the external search? If consumers have complete information and act rationally, they will choose brand j. However, in the real world, there are many alternatives, and it is difficult to choose the best brand based on a complete evaluation of alternatives.

Of the two forecasting methods established in section 5.1, method 1 assumes that the purchase probability of brands that are not members of the experiential set is 0, while method 2 estimates the probability of brands outside the set from their hierarchical structures (demographic variables) and the external search probability. As a result, method 2 outperforms method 1. This means that when consumers conduct an external search, they are able to choose their preferred brands to some degree. Because of that, if their external searches work ineffectively, the predictive accuracy of method 2 would be the same as that of method 1. Although this result also implies that the brand communication of firms is adequate, we are able to say that consumers' external searches work effectively to a certain degree.

From these predictions, we can confirm that the external searches of consumers work well. Firms also use the model more actively such as for recommending preferred brands (e.g., Ansari, Essegaier and Kohli, 2000;

Ansari and Mela, 2005). Because many previous recommendation models have estimated consumer preferences from the purchase records of others, they risk recommending a brand that is purchased but does not satisfy consumer needs. By contrast, the proposed model estimates the repeat purchase intentions of customers for a particular brand and thus reduces the above risk. The proposed model is therefore theoretically better because firms can offer brands that will satisfy customers' needs.

6.2. Size of the Experiential Set

In this research, we use the first 9 months of the overall study period as the formation period and analyze the purchase behavior and dynamic change of the experiential set over the following 15 months.

At the end of the observation period, the average size of the experiential set is 2.25 and the maximum is 8. Although this research focuses on analyzing the top 10 brands, no one customer purchased all 10 brands. A total of 108 consumers (27%) purchased only one brand; therefore, the size of the experiential set of these consumers is 1. Furthermore, 120 consumers (30%) had two brands in their experiential sets and 76 consumers (19%) had three brands. In summary, 76% of consumers purchase fewer than or equal to three brands. We also find that many consumers tend to purchase the same brand. Hauser and Wernerfelt (1990) report that the size of the consideration set of the shampoo category is 6.1. Compared with this value, the size of the experiential set obtained in this research is relatively small. Thus, it is possible that some brands were considered but never purchased in the consideration set.

According to the finding that many consumers purchase only a limited selection of brands, it is difficult to obtain the market competitive structure (variance-covariance matrix) Σ straightforwardly. However, this research allows us to obtain the whole structure from the partial choice behavior of each consumer using the MCMC method.

Additionally, we have to take account of the upper limit of the experiential set. The proposed model does not assume that certain brands "drop out" of the long-term memory. This assumption is based on the proposition of Bettman (1979), who defines long-term memory as "permanent" and an "essentially unlimited store" (p. 151, Proposition 6.4).

This means that we do not have to consider the size limit of the set. Furthermore, most previous papers that have discussed the consideration set have not assumed that products drop out of this set. For example, Chiang, Chib, and Nrashimhan (1999) analyze the tomato ketchup market by assuming that the consideration set remains unchanged over 18 months. If the period of analysis is short, the lack of a so-called “dropout mechanism” would not cause a serious problem. However, for longer-term analysis we need to restrict the upper limit of the set or introduce a dropout mechanism.

7. Conclusion

In this research, we use scanner panel data to construct a stochastic brand choice model of consumer goods in which consumers repeatedly choose a brand from many alternatives. We then reexamine consumers’ repeat purchase behavior from the perspective of information processing theory.

This research makes three main contributions. First, we construct a theoretical framework to analyze behavioral data such as point-of-sales records with customer ID numbers. We also reexamine the concepts of internal search, external search, and learning proposed in the field of consumer studies. Furthermore, we reconstruct consumers’ repeat purchase behavior from the perspective of information processing theory. By introducing these concepts into the quantitative model, the proposed model is more theoretically valid. Furthermore, to define the experiential set that can be observed from purchase records, instead of from choice subsets such as the “consideration set” or “processing set,” we proposed a more practicable model. The proposed model is insusceptible to increases in the number of alternatives and is applicable even for markets comprising dozens of alternatives.

The second contribution is that we construct a high performing forecasting model. From the results derived using the validation set, we find that the proposed model has a high predictive ability. Because the model is designed so that the purchase probability of brands that are members of the experiential set is higher, this result implies that many consumers tend to choose brands that they have always purchased. This finding is in line with previous research. The model also predicts external searches with a high degree of accuracy. This implies that it is able to discriminate between consumers who tend to conduct external searches (i.e., “variety seekers”) and those that do not. This model also shows the attitude of each consumer for the focal product category.

The third contribution is that the proposed model has the potential to be flexible in terms of application and extension. Because most proposed model structures are based on previous brand choice models and are estimated using the MCMC method, we can easily incorporate the specific model structure developed in previous research. We would thus be able to extend the model into, for example, a cross-category or dynamic (time-series) model.

For future research, we highlight the following two issues. First, future studies should examine the proposed model using other product categories. Although we have found a high level of predictive accuracy by focusing on the shampoo category, we must confirm that the proposed model has such a predictive ability in other categories. This research shows the validity of the concept of the experiential set as a result of predictions; however, we need to report the existence of the experiential set in other products. The second issue is the reconsideration of the model’s assumptions. As discussed in section 6, it is desirable to incorporate a dropout mechanism. When the period of analysis is extended, such a dropout mechanism will become more important. Future research should thus reexamine other parts of the model as the need arises.

A. Detailed Description of the Model

In this section, we will show the detailed description and estimation procedure of the model. At first, recall the purchase probability of consumer i's t-th purchase occasion of brand j, which is a member of the experiential set E_{it}. The proposed model is defined as follows:

$$y_{itj} = \begin{cases} 1 & \text{if } y^*_{itj} > 0 \\ 0 & \text{if } y^*_{itj} \leq 0 \end{cases}$$
$$y^*_{itj} = x_{it}\beta_{ij} + \varepsilon_{itj}, \quad j \in E_{it} \tag{A.1}$$

To consider the interaction of each brand, let $\tilde{y}^*_{it}$ be a $n(E_{it})$-dimensional vector consisting of the elements in $y^*_{it} = \{y^*_{it1}, \cdots, y^*_{itJ}\}'$ that correspond to the brands contained in E_{it}. In addition, left $\widetilde{B}^*_{it}$ be a $n(E_{it}) \times K$ matrix consisting of the elements in $B_i = \{\beta'_{i1}, \cdots, \beta'_{iJ}\}'$ that correspond to E_{it}. $\widetilde{\Sigma}$ is a $n(E_{it}) \times n(E_{it})$ partial matrix reconstructed from the variance-covariance matrix Σ. With these expressions, the following multivariate regression model can be formed:

$$\tilde{y}^*_{it} = \widetilde{B}_{it}x_{it} + \tilde{\varepsilon}_{it}, \quad \tilde{\varepsilon}_{it} \sim N(0, \tilde{\Sigma}) \tag{A.2}$$

The external search behavior of consumer i on the t-th purchase occasion is explained by following regression model:

$$z_{it} = \begin{cases} 1 & \text{if } z^*_{it} > 0 \\ 0 & \text{if } z^*_{it} \leq 0 \end{cases}$$
$$z^*_{it} = w_{it}\gamma_i + \zeta_{it} \tag{A.3}$$

where we assume, in this study, that $x_{ij} = w_{ij}$.

A hierarchical structure to explain b_i and γ_i is defined as follows (where $\theta_i = \{b'_i, \gamma'_i\}'$):

$$\theta_i = \Delta v_i + \xi_i, \qquad \xi_i \sim N(0, \Lambda) \tag{A.4}$$

A.1.Prior distributions

At first, we assume the prior distribution of Σ based on Manchanda, Ansari, and Gupta (1999):

$$\pi(\Sigma) \propto \exp\left\{-\frac{1}{2}\left(\mathrm{vec}^*(\Sigma) - \mathrm{vec}^*(\Sigma_0)\right)\mathrm{G}_0\left(\mathrm{vec}^*(\Sigma) - \mathrm{vec}^*(\Sigma_0)\right)\right\} \qquad \text{(A.5)}$$

where $\mathrm{vec}^*(\mathrm{M})$ is a vector operator, which arranges the upper triangle elements of M in a vector. Therefore, $\mathrm{vec}^*(\Sigma)$ is a $\mathrm{J}(\mathrm{J}-1)/2$ -dimensional vector. We assume the hyper-parameters Σ_0 and G_0 are, respectively, a $\mathrm{J} \times \mathrm{J}$ and $\mathrm{J}(\mathrm{J}-1)/2$ identity matrix.

Since Δ is a $(\mathrm{J}+1)\mathrm{K} \times \mathrm{L}$ matrix, we assume the prior distribution is a matrix normal distribution that $\Delta \sim N(\Delta_0, \Omega_0, \Lambda)$, where Δ_0 is a $(\mathrm{J}+1)\mathrm{K} \times \mathrm{L}$ zero matrix and $\Omega_0 = 100 \times \mathrm{I}_{(\mathrm{J}+1)\mathrm{K}}$.

We also assume that Λ is a diagonal matrix and each element is a Gamma distribution. For stability of the estimate, let $\lambda_\mathrm{k} \sim G(\mathrm{q}_0/2,\ \mathrm{Q}_0/2)$, where $\mathrm{q}_0 = \mathrm{Q}_0 = 10^4$ for each element λ_k.

We decompose the equation of θ_i, assuming that Λ is a diagonal matrix.

$$\theta_\mathrm{i} = \begin{pmatrix}\mathrm{b_i}\\ \gamma_\mathrm{i}\end{pmatrix} = \begin{pmatrix}\Delta_\mathrm{b}\\ \Delta_\gamma\end{pmatrix}\mathrm{v_i} + \xi_\mathrm{i},\ \ \xi_\mathrm{i} \sim N\left(0, \begin{pmatrix}\Lambda_\mathrm{b} & 0\\ 0 & \Lambda_\gamma\end{pmatrix}\right) \qquad \text{(A.6)}$$

A.2. Posterior distributions

The full conditional posterior distribution of the model is defined using following functions. In this research, we introduce the latent variables based on Albert and Chib (1990) to solve the discrete choice model. At first, the brand choice term is defined as follows:

$$\mathrm{f}(\mathrm{y_i}, \mathrm{y_i^*}|\mathrm{B_i}, \Sigma) = \left\{\prod_{\mathrm{t}=1}^{\mathrm{T_i}}\left[\prod_{\mathrm{j}\in\mathrm{E_{it}}}\pi\left(\mathrm{y_{itj}}\middle|\mathrm{y^*_{itj}}\right)\right]\pi(\tilde{\mathrm{y}}^*_{\mathrm{it}}|\widetilde{\mathrm{B}}_{\mathrm{it}}, \tilde{\Sigma}_{\mathrm{it}})\right\} \qquad \text{(A.7)}$$

The external search term is defined as follows:

$$g(z_{it}, z_i^*|B_i, \Sigma) = \prod_{t=1}^{T_i} \pi(z_{it}|z_{it}^*)\pi(z_{it}^*|\gamma_i) \qquad \text{(A.8)}$$

Using $f(\cdot)$ and $g(\cdot)$ above, the full conditional posterior distribution is defined as follows (to simplify the expression, let ϑ be the set of parameters, D the set of data, and $\theta = \{b_i, \gamma_i\}$):.

$$\pi(\vartheta|D) \propto \left[\prod_{i=1}^{N} f(y_i, y_i^*|B_i, \Sigma) g(z_{it}, z_i^*|B_i, \Sigma)\pi(\theta|\Delta, \Lambda)\right]\pi(\Sigma)\pi(\Delta)\pi(\Lambda) \qquad \text{(A.9)}$$

We obtain the conditional posterior distributions from the above equation.

A. 2. 1. The conditional posterior of $\mathbf{y_{itj}^*}, \mathbf{i = 1, \cdots, N, t = 1, \cdots, T_i, j \in E_{it}}$:

The sample of the latent variable y_{itj}^* is drawn from a truncated normal distribution. There are some methods to obtain the random sample from a truncated normal distribution. Following Geweke (1991), we use different methods in this research, such as normal rejection sampling and the exponential rejection sampling (Devroye, 1986), depending on the threshold. Let $TN_{(a,b)}(m, s)$ be a truncated normal distribution with mean m and standard deviation s restricted to (a, b).

$$y_{itj}^*| \cdot \sim \begin{cases} TN_{(0,\infty)}(\mu_1, \sigma_1), & \text{if } y_{itj} = 1 \\ TN_{(-\infty,0]}(\mu_1, \sigma_1), & \text{if } y_{itj} = 0 \end{cases} \qquad \text{(A.10)}$$

where $\mu_1 = m_{itj} + S_{j,-j}S_{-j,-j}^{-1}(\tilde{y}_{it,-j} - m_{it,-j})$, $\sigma_1 = S_{jj} + S_{j,-j}S_{-j,-j}^{-1}S_{-j,j}$, $m_{ij} = \widetilde{B}_{it}x_{it}$, $S = \tilde{\Sigma}$.

A. 2. 2. The conditional posterior of $\mathbf{z_{it}^*}$, $\mathbf{i = 1, \cdots, N, t = 1, \cdots, T_i}$:

z_{it}^* is also drawn from the truncated normal distribution. If $z_{it} = 1$, similar to y_{itj}^*, $z_{it}^*| \cdot \sim TN_{(0,\infty)}(w_i'\gamma, 1)$. However, if $z_{it} = 0$, $z_{it}^*| \cdot \sim TN_{(-\infty,0]}(w_i'\gamma, 1)$.

A. 2. 3. The conditional posterior of $\mathbf{B_i}$, $\mathbf{i = 1, \cdots, N}$:

Since we cannot obtain a $J \times K$ matrix parameter B_i, we obtain a K-dimensional vector β_{ij} for J alternatives. From the marginal distribution of multivariate normal, an element of multivariate regression is expressed as follows:

$$\tilde{y}^{*}_{itk} = \beta'_{ik} x_{it} + e_{itk}, e_{itk} \sim N(\mu_{itk}, \sigma_{itk}) \tag{A.11}$$

where $\mu_{itk} = S_{k,-k} S^{-1}_{-k,-k} \left(\tilde{y}^{*}_{it,-k} - \widetilde{B}_{it,-k} x_{it}\right)$, $\sigma_{itk} = S_{kk} - S_{k,-k} S^{-1}_{-k,-k} S_{-k,k}$, and $S = \tilde{\Sigma}_{it}$. In the same manner, we can decompose the prior distribution. In this paper, we denote the decomposition of b_i into β_{ij} as follows:

$$b_i = \begin{pmatrix} \beta_{i1} \\ \vdots \\ \beta_{iJ} \end{pmatrix} = \begin{pmatrix} \Delta_1 \\ \vdots \\ \Delta_J \end{pmatrix} v_i + \xi_i, \xi_i \sim N\left(0, \begin{pmatrix} \Lambda_{11} & \cdots & \Lambda_{1J} \\ \vdots & \ddots & \vdots \\ \Lambda_{J1} & \cdots & \Lambda_{JJ} \end{pmatrix}\right) \tag{A.12}$$

We obtain following equation by marginalizing vector β_{ij}:

$$\beta_{ij} = \Delta_j z_i + \zeta_{ij}, \zeta_{ij} \sim N(c_{ij}, L_{ij}) \tag{A.13}$$

where $c_{ij} = \Lambda_{j,-j} \Lambda^{-1}_{-j,-j} \left(b_{i,-j} - \Delta_{i,-j} z_i\right)$ and $L_{itk} = \Lambda_{jj} - \Lambda_{j,-j} \Lambda^{-1}_{-j,-j} \Lambda_{j,-j}$. However, in this paper, since we defined Λ as a diagonal, $c_{ij} = 0$, and $L_{itk} = \Lambda_{jj}$.

From equations (A.11) and (A.13), we obtain following posterior distribution:

$$\beta_{ij} | \cdot \sim N(m_1, V_1) \tag{A.14}$$

where $V_1 = \left(\sum_{t=1}^{T_i} I_{itj} \sigma^{-1}_{itj} x'_{it} x_{it} + L^{-1}_{ij} \left(\Delta_j z_i + c_{ij}\right)\right)^{-1}$, and $m_1 = V_1 \left(\sum_{t=1}^{T_i} I_{itj} \sigma^{-1}_{itj} x'_{it} \left(y^{*}_{itj} - \mu_{itj}\right) + L^{-1}_{ij} \left(\Delta_j z_i + c_{ij}\right)\right)$. Additionally, I_{itj} is an indicator function defined as $I_{itj} = \mathbf{1}(j \in E_{it})$.

A. 2. 4. The conditional posterior of $\boldsymbol{\Sigma}$:

As mentioned above, a $J \times J$ matrix parameter Σ expresses the competitive structure of the whole market. We must obtain the whole matrix

Σ, although in the individual model we use the partial matrix $\tilde{\Sigma}$, which consists of the elements of Σ that correspond to the brands contained in E_{it}. However, all diagonal elements of Σ must be 1, and the matrix needs to be positive definite (Chib and Greenberg, 1998; Manchanda, Ansari, and Gupta, 1999). Since if Σ is a positive definite matrix, we can assure that the partial matrix $\tilde{\Sigma}$ is also positive definite, we only need to construct an appropriate matrix Σ.

We must obtain the posterior distribution of Σ from the likelihood of y_{it}^* and the prior distribution of Σ. We can express the posterior as a product of these densities:

$$\pi(\Sigma|\cdot) \propto \left\{ \prod_{i=1}^{N} \prod_{t=1}^{T_i} \pi\left(\tilde{y}_{it}^* \middle| \tilde{B}_{it}, \tilde{\Sigma}_{it}\right) \right\} \times \pi(\Sigma) \tag{A.15}$$

However, there is no well-known distribution to satisfy the restriction of Σ. Authors such as Edwards and Allenby (2003), Chib and Greenberg (1998), and Manchanda, Ansari, and Gupta (1999) have proposed some methods to construct the matrix. In this research, we obtain the candidate sample and the candidate distribution based on Manchanda, Ansari, and Gupta (1999). Let the candidate distribution be $q(\Sigma)$ and the candidate sample Σ^c. Then, the acceptance rate a is obtained from the following equation:

$$a = \min\left\{ \frac{\pi(\Sigma^c|\cdot)/q(\Sigma^c)}{\pi(\Sigma|\cdot)/q(\Sigma)}, 1 \right\} \tag{A.16}$$

A. 2. 5. The conditional posterior of $\boldsymbol{\gamma}_i$, $\mathbf{i = 1, \cdots, N}$:

$$\gamma_i|\cdot \sim N(m_1, S_1) \tag{A.17}$$

where $S_1 = \left(\Lambda_\gamma^{-1} + W_i' W_i\right)^{-1}$, $m_1 = S_1\left(\Lambda_\gamma^{-1}\Delta_\gamma v_i + W_i' z_i^*\right)$, $W_i = \left(w_{i1}, \cdots, w_{i,T_i}\right)'$,

and $z_i^* = \left(z_{i1}^*, \cdots, z_{i,T_i}^*\right)'$.

A. 2. 6. The conditional posterior of $\boldsymbol{\Delta}$:

$$\Delta | \cdot \sim N(\Delta_1, \Omega_1, \Lambda) \tag{A.18}$$

where $\Omega_1 = (\Omega_0^{-1} + V'V)^{-1}$, and $\Delta_1 = (\Theta'V + \Delta_0\Omega_0^{-1})\Omega_1$. Δ is a sample $(J+1)K \times L$ matrix normal distribution. Refer to Rowe (2002) and Dawid (1981) for further details of the matrix distribution.

A. 2. 7. The conditional posterior of $\boldsymbol{\lambda_k}, \mathbf{k} = \mathbf{1}, \cdots, (\mathbf{J}+\mathbf{1})\mathbf{K}$:

$$\lambda_k | \cdot \sim G(q_1/2, Q_1/2) \tag{A.19}$$

where $q_1 = q_0 + N$, $Q_1 = Q_0 + \sum_{i=1}^{N}(\theta_{ik} - \delta_{k\cdot}' v_i)^2 + \left(\delta_{k\cdot} - \delta_{0,k\cdot}\right)' \Omega_0^{-1}\left(\delta_{k\cdot} - \delta_{0,k\cdot}\right)$, $\delta_{k\cdot}$ is k-th column vector of Δ, and $\delta_{0,k\cdot}$ is k-th column vector of Δ_0. In both vectors, the number of elements is L.

A. 3. Initial values and sample collections

For initial values, we let all latent variables y_{itj}^* and z_{it}^* be 0, and parameters B_i, γ_i, and Δ be zero vectors (matrices). In addition, we let Σ be an identity matrix and all λ_k be 1.

We ran the chain for 15,000 iterations. The result was reported on a sample of 10,000 draws from the posterior distribution, after we discarded 5,000 burn-in draws.

References

Albert, J. H., and Chib, S. (1993). Bayesian Analysis of Binary and Polychotomous Response Data, *Journal of the American Statistical Association*, 88, 669-679.

Allenby, G. M., and Rossi, P. E.(1999). Marketing Models of Consumer Heterogeneity, *Journal of Econometircs*, 89, 57-78.

Andrews, R. and Srinivasan, T. C. (1995). Studying Consideration Effects in Empirical Choice Models Using Scanner Panel Data, *Journal of Marketing Research*, 32, 30-41.

Ansari, A., and Mela, C. F.(2003). E-Customization, *Journal of Marketing Research*, 40, 131-145.

Ansari, A., Essegaier, S., and Kohli, R. (2000). Internet Recommendation Systems, *Journal of Marketing Research*, 37, 363-375.

Bettman, J. R. (1970). Information Processing Models of Consumer Behavior, *Journal of Marketing Research*, 7, 370-376.

Bettman, J. R. (1971). The Structure of Consumer Choice Process, *Journal of Marketing Research*, 13, 465-471.

Bettman, J. R. (1979). An Information Processing Theory of Consumer Choice, Addison-Wesley Publishing Company.

Bettman, Johnson and Payne (1991) Consumer Decision Making, In T. S. Robertson & H. H. Kassarjian (Eds), *Handbook of consumer behavior (pp. 50-84)*, Prentice Hall.

Blackwell, R. D., Miniard, P. W., and Engel, J. F. (2006) *Consumer Behavior 10th ed.*, OH: Thomson Higher Education.

Blattberg, R. C., Kim, B., and Neslin, S. A. (2008). *Database Marketing: Analyzing and Managing Customers*, Springer.

Brisoux, J. E. and Cheron, E. J. (1990). Brand Categorization and Product Involvement, *Advances in Consumer Research*, 17, 101-109.

Chiang, J. (1991). A Simultaneous Approach to the whether, What and How Much to Buy Questions, *Marketing Science*, 10(4), 297-315.

Chiang, J., Chib, S., and Narasimhan, C. (1999). Markov Chain Monte Carlo and Models of Consideration Set and Parameter Heterogeneity, *Journal of Econometrics*, 89, 223-248

Chen, M. H., Shao, Q. M., and J. G. Ibrahim (2000). *Monte Carlo Methods in Bayesian Compu-tation*, New York: Springer.

Chib, S., and Greenberg, E. (1998). Analysis of Multivariate Probit Models, *Biometrika*, 85, 347-361.

Chib, S., Seetharaman, P. B., and Strijnev, A. (2004). Model of Brand Choice with a Non-Purchase Option Calibrated to Scanner-Panel Data, *Journal of Marketing Research*, 41(May), 184-196.

Chintagunta, P. K. (1993). Investigating Purchase Incidence, Brand Choice and Purchase Quantity Decisions of Households, *Marketing Science*, 12(2), 184-208.

Dawid, A. P. (1981). Some Matrix-Variate Distribution Theory: Notational Considerations and a Bayesian Application, *Biometrika*, 68, 265-274.

Devroye, L. (1986). *Non-Uniform Random Variable Generation*, New-York: Springer.

Edwards, Y. D., and Allenby, G. M. (2003). Multivariate Analysis of Multiple Response Data, *Journal of Marketing Research*, 40(3), 321-334.

Engel, J. F., Kollat, D. T., and Blackwell, R. D. (1968). *Consumer Behavior*, Rinehart & Winston.

Farley, J. U., and Ring, L. W. (1974). An Empirical Test of the Howard-Sheth Model of Buyer Behavior, *Journal of Marketing Research*, 7(Nov.), 427-438.

Fishbein, M. (1967). *A Consideration of Beliefs and Their Role in Attitude Measurement*, New York: Wiley.

Geweke, J. (1991). Efficient Simulation from the Multivariate Normal and Student-t Distributions Subject to Linear Constraints and the Evaluation of Constraint Probabilies, In E. Kermamidas (Ed.), *Proceedings of 23rd symposium on the interface between computing science and statistics (pp. 571-578)*.

Geweke, J. (1992). Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments, In J.O. Berger, J.M. Bernardo, A.P. Dawid, and A.F.M. Smith (Eds.) *Bayesian Statistics 4 (pp. 169-194)*, Oxford: Oxford University Press.

Gilbride, T. J. and Allenby, G. M. (2004). A Choice Model with Conjunctive, Disjunctive, and Compensatory Screening Rules, *Marketing Science*, 23(3), 391-406.

Guadani, P. M., and Little, J. D. C. (1983). A Logit Model of Brand Choice Calibrated on Scanner Data, *Marketing Science*, 2(3), 203-238.

Hauser, J. R., and Shugan, S. M. (1983). Defensive Marketing Strategies, *Marketing Science*, 2, 319-360.

Hauser, J. R., and Wernerfelt, B. (1990). An Evaluation Cost Model of Consideration Sets, *Journal of Consumer Research*, 16, 393-408.

Hensher, D., Louviere, J. and Swait, J. (1999). Combining Sources of Preference Data, *Journal of Econometrics*, 89, 197-221.

Howard, J. A., and Sheth, J. N. (1969). *The Theory of Buyer Behavior*, John Wiley & Sons.

Howard, J. A., Shay, R. P., and Green, C. A. (1988). Measuring the Effect of Marketing Information on Buying Intentions, *The Journal of Consumer Marketing*, 5(3), 5-14.

Hoyer, W. D., and MacInnis, D. J. (2008). *Consumer Behavior 5th ed.* South-Western Cengage Learning

Lapersonne, E., Laurent, G., and Le Goff, J-J. (1995). Consideration Sets of Size One: An Empirical Investigation of Automobile Purchases, *International Journal of Research in Marketing*, 12, 55-66.

Manchanda, P., Ansari, A., and Gupta, S. (1999). The “Shopping Basket”: a Model for Multicategory Purchase Incidence Decisions, *Marketing Science,* 18, 95-114.

McCulloch, R. E., and Rossi, P. E. (1994). An Exact Likelihood Analysis of the Multinomial Probit Model, *Journal of Econometrics*, 64, 207-240.

McCulloch, R. E., Polson, N. G., and Rossi, P. E. (2000). A Bayesian Analysis of the Multinomial Probit model with Fully Identified Parameters, *Journal of Econometrics*, 99, 173-193.

Mitchell, A. A. (1981). The Dimensions of Advertising Involvement, *Advance in Consumer Research*, 8, 25-30.

Mowen, J. C. (1995). *Consumer Behavior 4th ed.*, Macmillan Publishing Company.

Narayana, C. L., and Markin, R. J. (1975). Consumer Behavior and Product Performance: An Alternative Conceptualization, *Journal of Marketing*, 39, 1-6.

Nicosia, F. M. (1966). *Consumer Decision Process*, Prentice-Hall.

Nielop, E. V., Bronnenberg, B., Paap, R., Wedel, M., and Franses, P. H. (2010) Retrieving Unobserved Consideration Sets from Household Panel Data, *Journal of Marketing Research*, 47, 63-74

Roberts, J. H. (1989). A Grounded Model of Consideration Set Size and Composition, *Advances in Consumer Research*, 16, 749-757.

Roberts, J. H., and Lattin, J. M. (1991). Development and Testing of a Model of Consideration Set Composition, *Journal of Marketing Research*, 28, 429-440.

Rowe, D. B. (2002). *Multivariate Bayesian Statistics*, Chapman & Hall.

Shocker, A. D., Ben-Akiva, M., Boccara, B., Nedungadi, P. (1991). Consideration Set Influences on Consumer Decision-Making and Choice: Issues, Models, and Suggestions, *Marketing Letters*, 3, 181-197.

Silk, A. J., and Urban, G. L. (1978). Pre-Test Market Evaluation of New Packaged Goods: A Model and Measurement Methodology, *Journal of Marketing Research*, 15(2), 171-191.

Simon, H. A. (1947). *Administrative Behavior: A Study of Decision-Making Process in Administrative Organization*, Macmillan: New York.

Simon, H. A. (1997). *Administrative Behavior: A Study of Decision-Making Process in Administrative Organization 4th edition*, Free Press: New York.

Tanner, M. A., and Wong, W. H. (1987). The Calculation of Posterior Distributions by Data Augmentation, *Journal of the American Statistical Association*, 82, 528-540.

Train, K. E. (2003). Discrete Choice Methods with Simulation, Cambridge University Press.

Van Heerde, H., and Neslin, S. A. (2008). Sales Promotion Models, In B. Wierenga (Ed.) *Handbook of Marketing Decision Models (pp. 107-162)*, Springer.

Wright, P., and Barbour, F. (1977). Phased Decision Strategies: Sequels to Initial Screening, In Starr, M., and M. Zeleny (Eds.), *Multiple Criteria Decision Making. North-Holland TIMES Studies in Management Science*, Amsterdam: North-Holland.