MMRC

DISCUSSION PAPER SERIES

No. 413

A Brand Choice Model for Scanner Panel Data Using the Experiential Set

Sotaro Katsumata

Faculty of Economics, Nagasaki University

Makoto Abe

Graduate School of Economics, The University of Tokyo

August 2012

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

http://merc.e.u-tokyo.ac.jp/mmrc/dp/index.html

A Brand Choice Model for Scanner Panel Data

Using the Experiential Set

Sotaro Katsumata
Faculty of Economics, Nagasaki University
katsmat@nagasaki-u.ac.jp

Makoto Abe
Graduate School of Economics, The University of Tokyo
abe@e.u-tokyo.ac.jp

Abstract: In this study, 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 examine consumers' repeat purchase behavior from the perspective of information processing theory. In particular, we explicitly incorporate the concepts 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 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 $2^{\mathrm{n}} - 1$. 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 (Stimulus-Organism-Response) 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 (1970, 1971, 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*

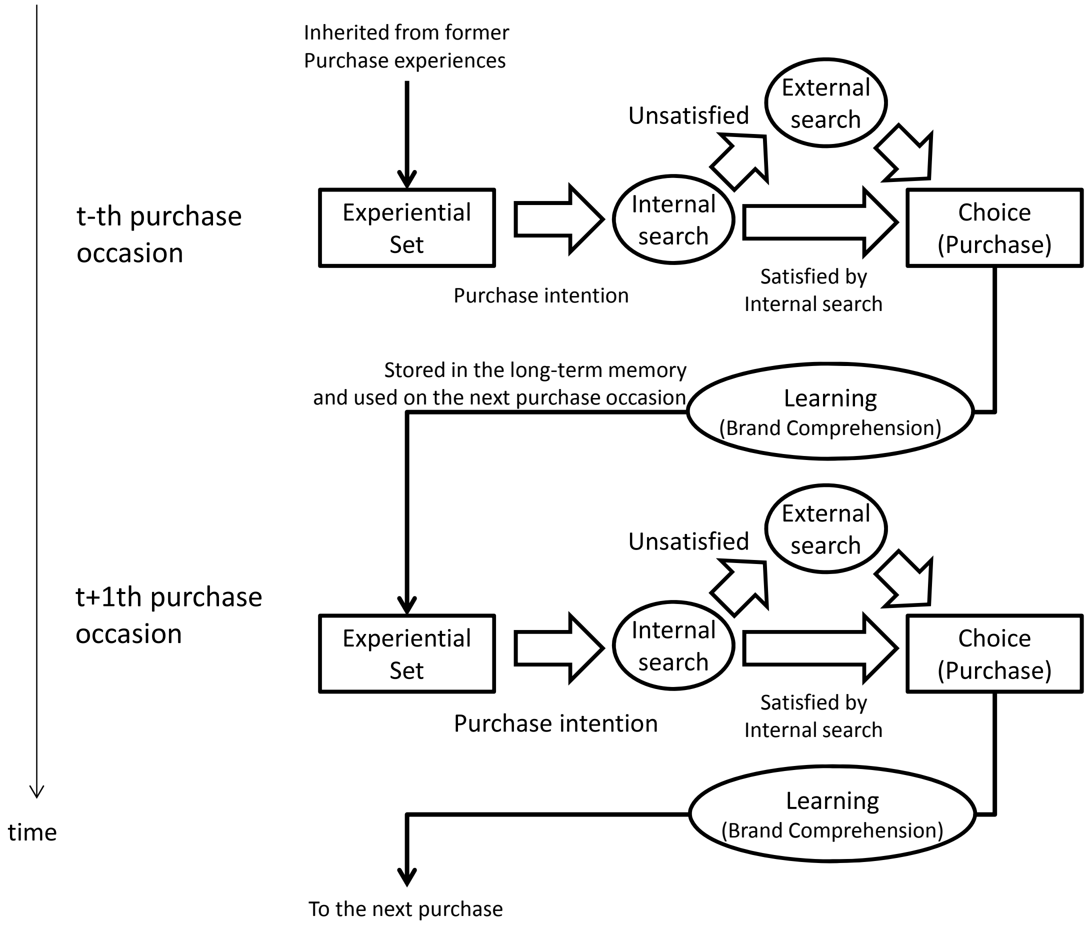
In this section, we examine how to incorporate the feedback system into a brand choice model. Because we cannot observe consumers' inner information processing decision from purchase data, we must somehow infer this search and learning.

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 Choice 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, $E_{i,t+1} = E_{it}$. In 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 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 be given as $y_{itj} = 1$. In this case, brand j is a member of the experiential set ($j \in E_{it}$); the purchase output variable of brand k, which is a member of E_{it} and is not purchased, is given 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 is given as $y_{it0} = 1$. Because brand j is not a member of set E_{it}, we do not use this occasion to estimate the purchase probability of brand j. As described in the definition of E_{it}, this experiential set is the set of brands that have already been purchased and used. Therefore, E_{it} expands not at the point of purchase but after the purchase.

3.2. The Choice Model for the Multinomial Observations

Before the model formulation, we show the basic framework of the multinomial choice model. In it, if there are J alternatives available, the most preferred alternative is chosen. Let the choice result observation of consumer i at t-th choise occasion be $y_{it1}, \cdots, y_{itJ}$, then any one element takes 1 (chosen) and other elements take 0 (not chosen). Then, we define the latent variables of evaluation $y^*_{it1}, \cdots, y^*_{itJ}$, which have following conditions: if $y_{itj} = 1$ then $y^*_{itj} = \max\{y^*_{it1}, \cdots, y^*_{itJ}\}$, and if $y_{itj} = 0$ then $y^*_{itj} < \max\{y^*_{it1}, \cdots, y^*_{itJ}\}$. This means that an observation will change depending on whether the evaluation is the highest or not. In practice, we have to set one alternative as the base alternative and fix the evaluation variable as a constant to satisfy the identification condition. For example, when we fix the evaluation of J-th alternative as 0, all of the evaluations of the alternatives other than the J-th alternative become a relative value, or $\{y^*_{it1} - y^*_{itJ}, \cdots, y^*_{it,j-1} - y^*_{itJ}, 0\}$. Corresponding parameters and variables that explain the evaluation will also be a relative value (e.g., Train, 2003). We also have to consider the identification condition of the variance-covariance matrix. We have to set a constraint to deal with the scale identification problem. We usually, set the (1,1) element of variance-covariance matrix Σ as 1 to satisfy the identification.

We now discuss these identification problems. We must first consider which alternative would be an appropriate base. To construct a hierarchical model that assumes consumer heterogeneity, the base alternative must be in the experiential set of all consumers for all periods. Though we have to estimate the parameters relative to the base alternative, we cannot find one. Therefore, we treat the external search y_{it0} as one of the alternatives and set this as the base alternative instead of any of the actual alternatives, such as concrete brands. Thus, the evaluation y^*_{itj} indicates the difference of evaluation between alternative j and the external search. For alternative j in the experiential set E_{it}, the relation between observation and evaluation is as follows:

$$y_{itj} = \begin{cases} 1 \text{ if } y^*_{itj} \geq \max\{(y^*_{itk} | k \in E_{it}), 0\} \\ 0 \text{ if } y^*_{itj} < \max\{(y^*_{itk} | k \in E_{it}), 0\} \end{cases}, \text{if } j \in E_{it} \qquad (1)$$

In a typical multinomial probit model, if there are J available alternatives, we have to set one as the base alternative and estimate the evaluations of the remaining $J-1$ alternatives. In this model, however, since we treat the external search as the $J+1$-th alternative and set this as the base, we estimate the evaluations of J alternatives.

It is consistent with the above theoretical discussions to set the external search as the base. When an individual evaluates the alternatives through an internal search, the preferable alternative will be chosen only if the alternative exceeds a minimum level of satisfaction. This minimum level of satisfaction is the threshold of the external search that is set as the base alternative. If an evaluation of all alternatives obtained from the internal search does not achieve the satisfaction level, the individual will find the new alternative through an external search. In this case, the evaluation of the new alternative will be higher than that of all the alternatives in the experiential set

Table 1: Correspondence relationship between observations and evaluations

choice occasion	observation(y)				evaluation(y^*)				Experiential set
	y_1	y_2	y_3	ES	y_1^*	y_2^*	y_3^*	ES	
0	-	-	-	0	-	-	-	0	{1}
1	1	0	0	0	> 0	-	-	0	{1}
2	1	0	0	0	> 0	-	-	0	{1}
3	0	1	0	1	< 0	-	-	0	{1}
4	1	0	0	0	$> \{y_2^*, 0\}$	$< y_1^*$	-	0	{1,2}
5	0	1	0	0	$< y_2^*$	$> \{y_1^*, 0\}$	-	0	{1,2}
6	0	0	1	1	< 0	< 0	-	0	{1,2}
7	0	0	1	0	$< y_3^*$	$< y_3^*$	$> \{y_1^*, y_2^*, 0\}$	0	{1,2,3}
8	0	1	0	0	$< y_2^*$	$> \{y_1^*, y_3^*, 0\}$	$< y_2^*$	0	{1,2,3}
9	0	0	1	0	$< y_3^*$	$< y_3^*$	$> \{y_1^*, y_2^*, 0\}$	0	{1,2,3}
10	0	1	0	0	$< y_2^*$	$> \{y_1^*, y_3^*, 0\}$	$< y_2^*$	0	{1,2,3}

Note: ES = external Search

Table 1 shows the process of choice subset formulation and brand evaluation. At the initial state, consumer's experiential set contains only brand 1. On the first and second purchase occasions, the consumer evaluates only brand 1 through internal search. As a result, the subsequent experiential set does not change. On the third purchase occasion, the consumer chooses brand 2, which is not in the experiential set. It implies that external search was conducted. The evaluation value of brand 1 for the third occasion is below the value of external search, which is 0. This means that the consumer is not satisfied with brand 1 and finds a new brand through external search. Similarly, on the sixth occasion, the consumer is not satisfied with the brands in the experiential set and chooses a new brand (brand 3) through external search. Brand 3 is now added to the experiential set, which will be evaluated through internal search at the seventh occasion.

Next, we define the structure of the variance-covariance matrix. This matrix explains the competitive or complement structure of the whole market. However, since the elements of the experiential set differ depending on individual and time, each consumer has only partial information. In this case, the variance-covariance matrix of a consumer becomes a partial matrix consisting of the alternatives in the experiential set. We have to estimate the whole matrix from this partial information. Since we need to estimate one matrix parameter for a whole market, we apply the same constraint as in the typical multinomial probit process, which fixes $(1,1)$ element as 1. We cannot obtain a random sample of the matrix parameter satisfying this constraint from any well-known distribution in a straightforward manner. However, simple approaches have been proposed (e.g., McCulloch and Rossi, 1994; Noible, 1998; McCulloch, Polson, and Rossi, 2000; Imai and van Dyk, 2005). For example, McCulloch, Polson, and Rossi, (2000) decompose the matrix into a vector and a partial matrix to obtain samples from a multivariate normal distribution and a Wishart distribution, respectively. Since we cannot apply these methods in this model, however, we obtain samples from a Gamma distribution for diagonal elements and apply the Metropolis-Hastings (M-H) method to obtain samples for the non-diagonal elements. Procedures are available for obtaining random samples of the variance-covariance matrix of a multivariate probit model using the M-H method, such as those in Chib and Greenberg (1998) and Manchanda, Ansari, and Gupta (1999). We have developed an estimation procedure based on these methods, a detailed description of which is provided in the Appendix.

We define the structure to explain the evaluation using y^*_{it} and Σ, as defined above. As is the multivariate probit experiential choice model, this model is constructed from a $n(E_{it})$-dimensional vector $\tilde{y}^*_{it}$ extracting a corresponding element from y^*_{it} and a $n(E_{it}) \times K$ parameter matrix $\widetilde{B}_{it}$, a $n(E_{it}) \times n(E_{it})$ parameter matrix $\tilde{\Sigma}_{it}$ extracting a corresponding element from B_i and Σ respectively.

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

We assume a hierarchical structure for parameter B_i. This prior structure supplies information using the customer-specific variable.

$$\mathrm{vec}(B_i) = \Delta w_i + \xi_i, \xi_i \sim N(0, V) \tag{3}$$

where w_i is a L-dimensional explanatory variable, Δ is a $JK \times L$ matrix parameter, and V is a $JK \times JK$ variance-covariance matrix. A detailed description of the prior and posterior distributions is provided in the Appendix.

4. Empirical Analysis

4.1. Data Overview and the Period of Analysis

We now conduct an empirical analysis of the model defined above. This section provides a data overview. We use the sales records for a drugstore chain’s laundry detergent provided by the *Joint Association Study Group of Management Science* and *Customer Communications*. The data period covers the two years from January 1, 2008 to December 31, 2009.

Since we need the formulation period of the experiential set, we use the first 9 months as the required period. The initial experiential set consists of the brands chosen during the formulation period, and the period of analysis thus spans the following 15 months. However, we use each consumer's last choice occasion for the forecasting set. We set the calibration period as this 15-month period except for the last purchase.

4.2. Studied Brands and Customers

For our studied brands, we choose the top 10 brands purchased within the study period; 300 customers were studied, as they had purchased these 10 brands more than four times during the calibration period.

4.3. Explanatory Variables

The proposed model has two kinds of explanatory variable, x_{it} and w_i. In x_{it}, we include the intercept, sale day dummy, and holiday dummy. In this store, all of the goods are discounted on the first and twentieth days of each month. Therefore, we will include a dummy variable for whether or not the day is a sale day. Similarly, we include a dummy variable for whether or not the day is a holiday. In w_i, we include the intercept, gender (female = 1), and logarithmic age.

4.4. Forecasting

To examine the forecasting ability of the model, we attempt a choice prediction for each customer's last choice occasion using the obtained parameter. We use the Hit Ratio and ROC scores as indicators of prediction performance. The Hit Ratio is the rate of customers who actually choose the brand predicted to be chosen. It is obtained from the prediction-observation matrix, which compares the forecast and the observation. The ROC score is usually used in the data mining performed by database marketing to examine the forecasting performance of a model obtained from the ROC curve. An ROC score is obtained for each brand. Further details may be found in Blattberg, Kim, and Neslin (2008).

5. Results

5.1. Forecasting Performances

This section describes the results of the model's forecast. First, we see the obtained ROC scores. We calculate two types of forecasting methods. In Method 1, the choice probabilities of the brands in the experiential set are obtained in the usual way, though the choice probabilities of the brands not in the experiential set are 0. In Method 2, the choice probabilities of the brands in the experiential set are the same as in Method 1. However, to obtain the choice probabilities of the brands not in the experiential set, one calculates the probabilities from estimated parameters and multiplies the external search probability to obtain the complement probabilities. Detailed procedures can be found in the Appendix.

Table 2 shows the obtained ROC scores of each brand. Each score has a value between 0 and 1. As predictive accuracy improves, the value approaches 1. If forecasting is completely random, the expectation value will be 0.5. We can thus say that, if the value exceeds 0.5, the model has some predictive ability. As shown in Table 2, the ROC scores of all brands exceed 0.5 substantially. We thus find that the predictive accuracy of the proposed model is fairly good.

Next, we see the Hit Ratio, defined as the number of consumers who choose the brand as predicted by the model divided by the total number of consumers. As with the ROC score, the Hit Ratio has a value between 0 and 1, and, as predictive accuracy improves, the value approaches 1. We calculate the value using the prediction-observation matrix (see Table 3). Table 3 displays choice forecasting as the column and actual choice as the row. The number of matching customers is expressed as the diagonal elements. We can obtain the Hit Ratio by dividing the diagonal elements by the total number.

Table 2: ROC Score

	ROC Score
Ariel	0.88
Attack	0.93
Bold	0.90
New Beads	0.83
Top	0.87
Blue Dia	0.93
PB (Private Brand)	0.91
FaFa	0.86
Liquid Top	NA
Style Fit	0.78

Note: we cannot obtain the result for Liquid Top because no one chose the brand on the last choice occasion

Table 3: Prediction and Observation Matrix

Pred.＼Obs.	Ariel	Attack	Bold	New Beads	Top	Blue Dia	PB	FaFa	Liquid Top	Style Fit	ES
Ariel	69	6	4	2	5	4	0	1	0	0	4
Attack	6	31	1	0	3	3	3	0	0	0	2
Bold	2	0	36	0	3	2	0	0	0	0	2
New Beads	2	1	5	12	6	7	0	0	0	0	4
Top	4	2	0	1	14	0	0	0	0	0	1
Blue Dia	1	0	2	1	3	21	0	0	0	0	0
PB (Private Brand)	0	2	1	0	1	4	4	1	0	0	3
FaFa	0	0	0	0	0	1	0	0	0	0	0
Liquid Top	0	0	0	0	0	0	0	0	0	0	0
Style Fit	2	0	0	0	0	1	0	0	0	1	1
External Search	1	0	0	0	0	0	0	0	0	1	0

The calculated Hit Ratio is 0.63, which we compare with other forecast values. If the forecasting is completely random, customers will have chosen any 10 brands or conducted an external search. The expectation value of the Hit Ratio is therefore $1/11 \approx 0.09$. The obtained Hit Ratio is substantially higher than this. Moreover, we could forecast that all customers will choose the brand with the largest share; the largest share brand, Ariel, was chosen by 87 customers, for a Hit Ratio of $87/300 \approx 0.29$. The Hit Ratio obtained by the proposed model also exceeds this value.

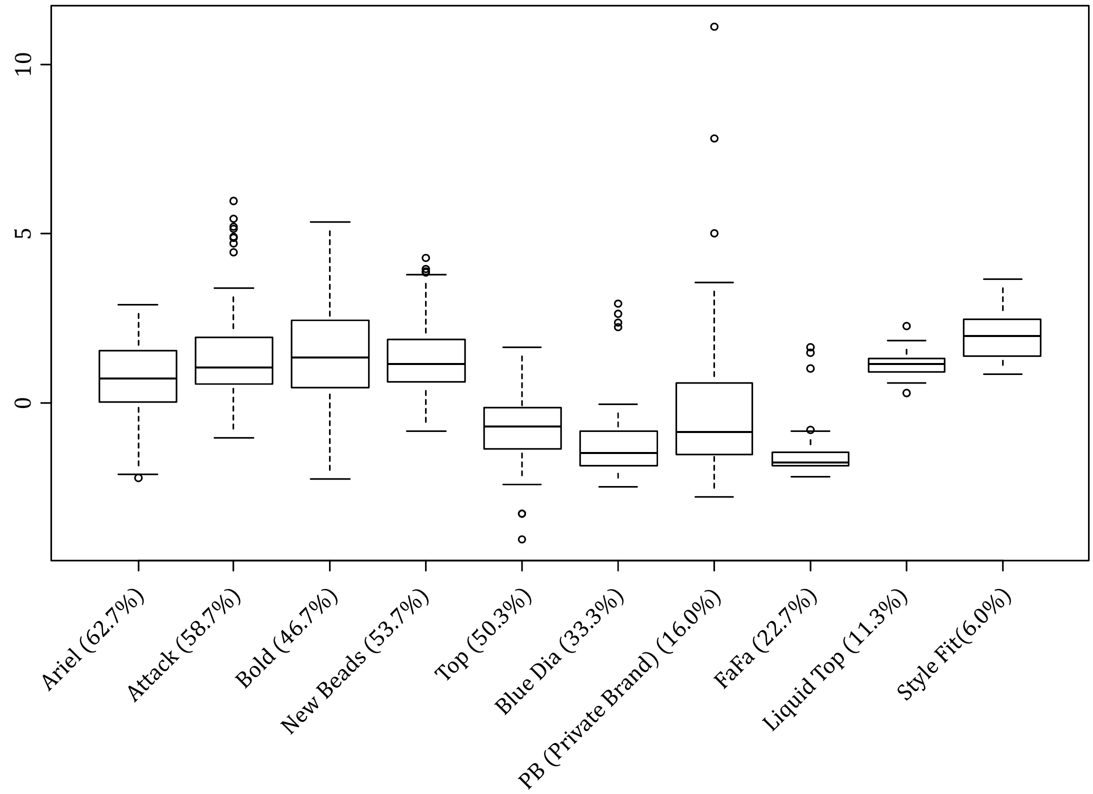
The proposed model is not only theoretically rigorous but also has high predictive ability, suggesting that it has practical applicability.

5.2. Parameters

In this section, we examine the model's parameters. We first show the distributions of the individual parameter β_{ij} for each brand. Figure 2 shows the parameters of each brand's intercept as the boxplot. The intercept of a brand contains each customer's basic preference. From this boxplot, we can find the distribution of the individual preferences for each brand. In the figure, the percentile value in the bracket is the rate of consumers who included the brand in their experiential set at the end of the calibration period. Note that, although we estimate the parameters of all the brands for all consumers, we include only the parameters of those consumers who include each brand in the experiential set in the figure.

We find differences within the variances for each brand. For example, few consumers highly preferred PB (Private Label), and BP's preference median (the middle line in the box) is not very high. Moreover, while the medians of Bold and New Beads are roughly equal, the distribution of Bold is more varied than that of New Beads.

Fig. 2: Brand intercept β_{ij1} of each consumer



Next, we show the parameters Δ and Σ. Table 4 presents the posterior mean of the selected elements of parameter Δ. We see the rough relation between each brand preference and demographic variable. 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 were calculated using the method proposed by Chen, Shao, and Ibrahim (2000). Table 5 presents the posterior mean of matrix Σ. In Table 4, we see two brands, the largest share brands Ariel and PB, and their differences. In the Sales-Gender parameter, Ariel is positive and PB is negative. On sale day, then, Ariel tends to be chosen by female consumers, while PB tends to be chosen by males. The table also shows that PB tends to be chosen by males during the holidays.

Table 4: Estimated Δ

Δ		Intercept	Gender	Age
Ariel	Intercept	-4.88	2.56	0.82
	Sale	-1.91	2.49 **	-0.21
	Holiday	-0.04	-0.73	0.23
PB (Private Brand)	Intercept	-6.13	-8.14	3.52
	Sale	5.19	-7.04 **	0.73
	Holiday	8.89	-3.95 **	-1.18

Table 5: Estimated Σ

Σ	Ariel	Attack	Bold	New Beads	Top	Blue Dia
Ariel	1.00					
Attack	-0.01	2.93				
Bold	0.01	0.02	3.63			
New Beads	0.00	0.02	0.02	3.83		
Top	0.00	0.02	0.02	0.00	5.05	
Blue Dia	0.02	0.01	0.02	-0.01	-0.03	5.79
PB (Private Brand)	0.03	0.02	-0.03	-0.02	-0.11	-0.08
FaFa	0.00	0.07	-0.02	-0.02	-0.04	-0.02
Liquid Top	0.03	-0.04	0.02	-0.12	-0.05	-0.38
Style Fit	-0.01	0.10	0.15	0.23	-0.32	-0.79 *

	PB	FaFa	Liquid Top	Style Fit
PB (Private Brand)	3.05			
FaFa	0.07	3.57		
Liquid Top	0.25	0.00	3.58	
Style Fit	0.06	-0.32	0.48	6.86

6. Discussion

6.1. Reject Brands in the Consideration Stage

Brisoux and Cheron (1990) propose a model in which the size of the consumer choice subset gradually narrows (known as the "Brisoux and Laroche Conceptualization"); this allows us to discuss the place of the experiential set in our model. We define the alternatives in the experiential set as the subset in which elements are evaluated through the internal search. Therefore, the experiential set is included in the processed set (originally, processed brands). The processed set, however, also includes the alternatives evaluated and processed through the external search. Strictly speaking, then, "the experiential set $\subseteq$ processed set." In addition, the experiential set is defined as the result of user experiences, formed though the feedback of user memories. This is positioned at the next step of the choice (preference stage) in the Brisoux and Laroche conceptualization. Since the model describes one choice occasion, the feedback system is not included. However, the experiential set is formed through the repeat choice process. In the processed set, we see that the experiential set contains the evoked, hold, and reject sets at the consideration stage. We will discuss these subsets of the obtained choice probabilities.

We can calculate the choice probabilities of the alternatives in the experiential set from the proposed model. Since the probabilities have values between 0 and 1, we cannot classify these three subsets precisely. However, we can classify them by referring to indicators. For example, we can obtain the external search probability for each consumer, which will be a preference threshold. If those alternative choice probabilities are lower than the external search probability, the preference for the alternatives will be very low.

Table 6 presents the choice probabilities for each studied consumer on the last choice occasion. Note that the choice probabilities were calculated only from the alternatives in the experiential set. The average choice probabilities of the first were obtained from the choice probabilities of the alternatives with the highest choice probabilities among the alternatives in the experiential set. The average choice probabilities of the first were obtained from the choice probabilities of the most preferred alternatives in the experiential set of each consumer. Likewise, the average choice probabilities of the second were obtained from choice probabilities of second preferred alternatives for each consumer. The table below shows the choice probabilities' degree of concentration.

Table 6: Choice Structure in the Consideration Stage

Size of the experiential set	Number of consumers	Average choice probabilities 1st	2nd	3rd	4th	below 5th	External search probability
1	38	0.970	-	-	-	-	0.030
2	52	0.788	0.176	-	-	-	0.036
3	62	0.705	0.205	0.058	-	-	0.031
4	54	0.640	0.216	0.074	0.028	-	0.043
5	42	0.552	0.235	0.102	0.050	0.012	0.061
over 6	52	0.520	0.208	0.115	0.058	0.023	0.077

We find that the probabilities of most of the preferred alternatives are very high, while the probabilities of the second most preferred alternatives are significantly lower. The choice probabilities of the third alternatives are even lower, roughly between 5 and 10 percent. Therefore, many consumers have only one preferred alternative and will not actively choose others, while the choice probabilities lower than the fourth most preferred alternatives are smaller than the external search probability. These alternatives may thus be classified as the reject set.

Therefore, consumers therefore have the following general inner structure in their choice of laundry detergent: they have one preferred alternative (which may be included in the evoked set), two or three comparable alternatives (which may be included in the hold set), and they rarely choose a third (or lower) alternative.

6.2. *Customer Satisfaction and Parameters*

We can draw implications for firm communication strategies from this experiential model. Our study divided consumers' first brand choice from their successive brand choice occasions because we wished to analyze the differences in evaluation processes between the first choice and subsequent occasions. For their first choice, consumers will evaluate the alternatives based on their expectations, while, after making this choice and having learned, they will evaluate based on their experience.

The difference between pre- and post-purchase evaluations has been discussed in customer satisfaction studies such as Oliver (2010). Theoretically, consumer satisfaction is affected by the disconfirmation between the performance expectations at pre-purchase and the perceived performance at post-purchase. If the expectation is too high, the consumer feels unsatisfied. It is not desirable to sell products to consumers who will not be satisfied by them through improper marketing communications that raise their expectations. Such a brand may not only be classified into the reject set discussed above and never again chosen by the consumer but may also inspire the consumer to generate negative word-of-mouth. A brand classified by many consumers into the reject set may present marketing communication challenges.

In this section, then, we will examine the relationship among three indicators defined below.

The first indicator, "experienced rate" is defined as the number of consumers who included the focal brand in their experiential set on the last choice occasion ($t = T_i$) divided by the total number of consumers. We define $n_j = \sum_{i=1}^{N} 1\left(j \in n\left(E_{i,T_i}\right)\right)$, where we omit the time subscript (applied hereafter), and the experienced rate of the alternative j is defined as n_j/N.

Second indicator, "rejected rate" is defined as the rate of experienced consumers who classify the focal brand as a reject set. Let us define the external search probability of consumer i as q_i and the choice probability of alternative j as p_{ij}; the number of rejected consumers is thus defined as $r_j = \sum_{(i,j)\in E_{it}} 1\left(p_{ij} < q_i\right)$. Using r_i, we define the rejected rate as r_j/n_j.

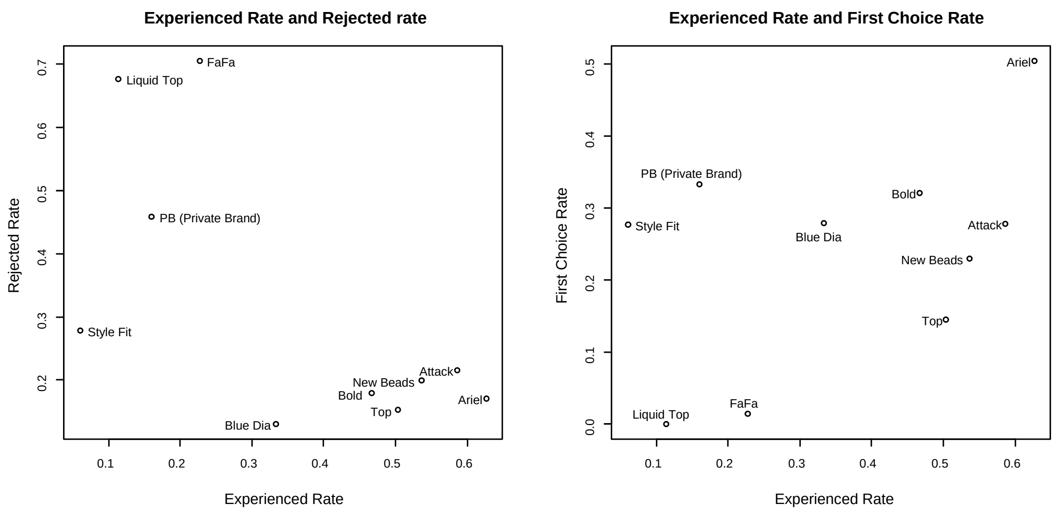
Finally, "first choice rate" is defined as the rate of consumers who prefer the focal brand. The first choice rate is defined as follows: f_j/n_j, where $f_j = \sum_{(i,j)\in E_{it}} \mathbf{1}\left(p_{ij} \geq \max\{pi\}\right.$, pi is a vector of the choice probabilities for the brands in the experiential set.

On the left side of Figure 4, we see that the rejected rates of many brands are roughly 0.3, while the rates of some brands (including PB) are higher than the others. Thus, though these brands were chosen once, they will not be chosen again. Though less than 20% of consumers purchased PB, almost half will not purchase it again.

On the right side of Figure 4, we find another aspect of the brands' relationship. Although the difference in experienced rates between Ariel and Attack is roughly 5%, the difference in the first choice rate is over 20%, and PB's first choice rate is the second highest. We can thus say that many people do not prefer PB but it has a substantial number of ardent fans.

We can draw useful implications from the result of the experiential set model that could guide a long-term brand strategy.

Fig. 4: Relationships between the experienced rate and the rejected or first choice rate



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.

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 laundry detergent 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 to incorporate a dropout mechanism of brands from an experiential set. When the duration of analysis is long, such a dropout mechanism will become more important. Future research should thus reexamine other parts of the model as the need arises.

A. Appendix

A.1. Prior Distributions and a Full Conditional Posterior Distribution

Recall the model proposed in this research. Let $\tilde{y}^*_{it}$ be an $n(E_{it})$-dimensional vector extracting corresponding elements from y^*_{it}. Let $\widetilde{B}_{it}$ be an $n(E_{it}) \times K$ partial matrix consisting of a $J \times K$ matrix B_i and $\tilde{\Sigma}_{it}$ an $n(E_{it}) \times n(E_{it})$ matrix parameter consisting of a $J \times J$ matrix Σ.

$$y_{itj} = \begin{cases} 1 \text{ if } y^*_{itj} \geq \max\{(y^*_{itk} | k \in E_{it}), 0\} \\ 0 \text{ if } y^*_{itj} < \max\{(y^*_{itk} | k \in E_{it}), 0\} \end{cases}, \text{if } j \in E_{it} \tag{A.1}$$

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

$$\text{vec}(B_i) = \Delta w_i + \xi_i, \xi_i \sim N(0, V) \tag{A.3}$$

We assume prior distributions for Σ, Δ, and $V = \text{diag}(v_1, \cdots, v_{JK})$. Let $\Delta \sim N(\Delta_0, V, S_{\Delta 0})$ and $v_k \sim G(\nu_0/2, V_0/2)$. To estimate Σ, we generate samples of its diagonal and off-diagonal elements separately. We define the diagonal elements of Σ as d, hence $d = \text{diag}(\Sigma)$. We define the off-diagonal lower triangle elements of Σ as s. To introduce the operator $\text{vec}^*(\cdot)$ that vectorizes the lower triangle elements of the objective matrices, we can denote that $s = \text{vec}^*(\Sigma)$. The prior distributions are then defined as follows:

$$d_j \sim IG\left(\frac{q_0}{2}, \frac{Q_0}{2}\right), j = 2, \cdots, J \tag{A.4}$$

$$\pi(s) \propto \exp\{-0.5(s - s_0)' G_0^{-1} (s - s_0)\} \tag{A.5}$$

where, $IG(a, b)$ is an Inverse-Gamma distribution with parameter $\{a, b\}$. We define a prior distribution of Σ from the above two distributions. To be a variance-covariance matrix of the discrete choice model, we need to introduce some restrictions for Σ. First, all diagonal elements of Σ have to be positive, and Σ needs to be a positive-definite matrix. Additionally, in MNP, we need a scale restriction that fixes the $(1,1)$ element as 1. In this research, we set the region satisfying the above restrictions as set C. Using set C, we define the prior distribution as follows:

$$\pi(\Sigma) \propto \prod_{j=2}^{J} \pi(d_j) \times \pi(s) \times 1(\Sigma \in C) \tag{A.6}$$

where $1(\cdot)$ is an indicator function. This expression is from Chib and Greenberg (1998), who propose an MVP estimation method.

From the above settings, the full conditional posterior distribution is defined as follows:

$$\pi(\theta|D) \\ \propto \left\{\prod_{i=1}^{N}\left[\prod_{t=1}^{T_i}\pi(\tilde{y}_{it}|\tilde{y}_{it}^{*})\pi\left(\tilde{y}_{it}^{*}\middle|\widetilde{B}_{it},\tilde{\Sigma}_{it}\right)\right]\pi(B_i|\Delta,V)\right\}\pi(\Delta|V)\pi(V)\pi(\Sigma) \tag{A.7}$$

where, θ is a set of parameters and D is a set of inputs. Moreover, $\pi(\tilde{y}_{it}|\tilde{y}_{it}^{*})$ is defined as follows (Albert and Chib, 1993):

$$\begin{aligned}\pi(\tilde{y}_{it}|\tilde{y}_{it}^{*}) = \prod_{j\in E_{it}} & 1\{(y_{itj}=1)\wedge(y_{itj}^{*}\geq \max(y_{it}^{*},0))\} \\ & + 1\{(y_{itj}=0)\wedge(y_{itj}^{*}\leq \max(y_{it}^{*},0))\}\end{aligned} \tag{A.8}$$

A.2. Posterior Distributions

A.2.1. Conditional Posterior of $\mathbf{y}_{itj}^{*}$

The sample of the posterior distribution of y_{itj}^{*} is obtained by following truncated normal distributions:[1]

$$y_{itj}^{*}|\cdot\sim\begin{cases}TN_{(a,\infty)}(\mu,s) & \text{if} \quad y_{itj}=1 \\ TN_{(-\infty,a]}(\mu,s) & \text{if} \quad y_{itj}=0\end{cases}, \quad \text{if} \quad j\in E_{it} \tag{A.9}$$

where, $a=\max\{(y_{itk}^{*}|k\in E_{it}),0\}$, $\mu=m_j+S_{j,-j}S_{-j,-j}^{-1}e_{-j}$, $s=\Sigma_{jj}-S_{j,-j}S_{-j,-j}^{-1}S_{-j,j}$, $S=\tilde{\Sigma}_{it}$, $m=\widetilde{B}_{it}x_{it}$, and $e_j=m_j-\tilde{y}_{itj}^{*}$. Note that subscript j is not a j-th element of the partial matrix/vector but aj-th element of the full matrix/vector.

[1] In practice, to stabilize the simulation, we introduced a restriction that sets the threshold maximum/minimum value. To introduce the threshold, let the interval of the TN be (a,C) and $(-C,a]$. Since this threshold C is attached to the prior distribution (A.8), the restriction will not cause any theoretical problems. In this research, we set $C=10$.

A.2.2. Conditional Posterior of Σ

The posterior sample of Σ is separately obtained from d_j and s, since we apply the random walk M-H to obtain s the same way as MVP; see Chib and Greenberg (1998) for further details. In this section, we will show a method of obtaining d_j.

The conditional posterior distribution of d_j is as follows:

$$\pi(d_j|\cdot) \propto \prod_{i=1}^{N}\prod_{t=1}^{T}|\tilde{\Sigma}_{it}|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}(\tilde{y}_{it}^{*}-\tilde{B}_{it}x_{it})'\tilde{\Sigma}_{it}^{-1}(\tilde{y}_{it}^{*}-\tilde{B}_{it}x_{it})\right\} \times \pi(d_j) \qquad (A.10)$$

We show the decomposition for the whole matrix Σ as a general case. First, using LDU decomposition, let us decompose the matrix into a lower triangle matrix L and a diagonal matrix D, where $D = \mathrm{diag}(d)$. From $\Sigma = LDL'$, we can decompose the determinant into $|\Sigma|^{-\frac{1}{2}} = |L'|^{-\frac{1}{2}}|D|^{-\frac{1}{2}}|L|^{-\frac{1}{2}}$. Additionally, since D is a diagonal matrix, we obtain $|D|^{-\frac{1}{2}} = |d_1|^{-\frac{1}{2}}\cdots|d_J|^{-\frac{1}{2}}$. Similarly, using the relation $\Sigma^{-1} = (LDL')^{-1} = (L')^{-1}D^{-1}L^{-1}$, we can rewrite the quadratic form. Let $L^{-1}(y_{it}^{*} - B_i x_{it}) = e_{it}$, we obtain $(y_{it}^{*} - B_i x_{it})'\Sigma^{-1}(y_{it}^{*} -$ Bixit=e′D−1e. Since D is a diagonal matrix, e′D−1e=e1d1−1e1+⋯+eJdJ−1eJ. From these relations, we can obtain the following equation:

$$\pi(d_j|\cdot) \propto \prod_{i=1}^{N}\prod_{t=1}^{T}\left[|d_j|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}e_j d_j^{-1} e_j\right\}\right]^{1(j\in E_{it})} \times \pi(d_j) \qquad (A.11)$$

Note that, in the t-th choice occasion of consumer i, e_j would not exist if j-th alternative would not be a member of E_{it}. Then, we operate this case introducing an indicator function $1(j \in E_{it})$. From this equation, we find that the posterior distribution of d_j follows a gamma distribution:

$$d_j|\cdot \sim IG\left(\frac{q_1}{2},\frac{Q_1}{2}\right), j = 2,\cdots,J \qquad (A.12)$$

where, $q_1 = \sum_{i=1}^{N}\sum_{t}^{T_i} 1(j \in E_{it}) + q_0$ and $Q_1 = \left(\sum_{i=1}^{N}\sum_{t=1}^{T_i} 1(j \in E_{it})e_j^2 + Q_0^{-1}\right)^{-1}$.

A.2.3. Conditional Posterior of $\mathbf{B_i}$

Since we cannot obtain a $J \times K$ matrix parameter B_i from a well-known distribution, we split up B_i into vectors and 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:

$$\tilde{y}^*_{itk} = \beta'_{ik} x_{it} + e_{itk}, e_{itk} \sim N(\mu_{itk}, \sigma_{itk}) \tag{A.13}$$

where $\mu_{itk} = S_{k,-k} S^{-1}_{-k,-k} \left(\tilde{y}^*_{it,-k} - \tilde{B}_{it,-k} x_{it}\right)$, $\sigma_{itk} = S_{kk} - S_{k,-k} S^{-1}_{-k,-k} S_{-k,k}$, 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:

$$b_i = \begin{pmatrix} \beta_{i1} \\ \vdots \\ \beta_{iJ} \end{pmatrix} = \begin{pmatrix} \Delta_1 \\ \vdots \\ \Delta_J \end{pmatrix} v_i + \xi_i, \xi_i \sim N\left(0, \begin{pmatrix} V_{11} & \cdots & V_{1J} \\ \vdots & \ddots & \vdots \\ V_{J1} & \cdots & V_{JJ} \end{pmatrix}\right) \tag{A.14}$$

where $V_{11} = \mathrm{diag}(v_1, \cdots v_K)$; therefore, for the j diagonal block, $V_{jj} = \mathrm{diag}\left(v_{(j-1)K+1}, \cdots, v_{jK}\right)$. Since we defined V as a diagonal, we obtain the following equation from the partial vector β_{ij}:

$$\beta_{ij} = \Delta_j z_i + \zeta_{ij}, \zeta_{ij} \sim N(0, V_{jj}) \tag{A.15}$$

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

$$\beta_{ij} | \cdot \sim N(m_1, V_{\beta 1}) \tag{A.16}$$

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

A.2.4. The Conditional Posterior of $\mathbf{\Delta}$ *and* $\mathbf{V}$

The posterior sample of Δ is obtained from following matrix normal distribution.

$$\Delta | \cdot \sim N(\Delta_1, V, S_{\Delta 1}) \tag{A.17}$$

where $S_{\Delta 1} = (S^{-1}_{\Delta 0} + W'W)^{-1}$, and $\Delta_1 = (\mathrm{vec}(B_i)'W + \Delta_0 S^{-1}_{\Delta 0}) S_{\Delta 1}$, $W = (w'_1, \cdots, w'_N)'$. Δ follows $JK \times L$ matrix normal distribution. Refer to Rowe (2002) and Dawid (1981) for further details of the matrix normal distribution.

The posterior sample of $v_k, k = 1, \cdots, JK$ is obtained from following gamma distribution.

$$v_k | \cdot \sim G(\nu_1/2, V_1/2) \tag{A.18}$$

where $\nu_1 = \nu_0 + N$, $V_1 = V_0 + \sum_{i=1}^{N} (b_{ik} - \delta'_{k\cdot} v_i)^2 + \left(\delta_{k\cdot} - \delta_{0,k\cdot}\right)' \Omega_0^{-1} \left(\delta_{k\cdot} - \delta_{0,k\cdot}\right)$, v_k is k-th element of vectorilized parameter $\mathrm{vec}(B_i)$, $\delta_{k\cdot}$ 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}^* be 0 and parameters B_i and Δ be zero vectors or zeros matrix. In addition, we let Σ be an identity matrix.

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.

A.4. Choice Probabilities

In this section, we will show the detailed discussion of choice probabilities.

First, the choice probability of an alternative in an experiential set is obtained from the following equation:

$$\Pr\left(y_{itj} = 1 \middle| j \in E_{it}\right) = \int_{a_1} \cdots \int_{a_{J^*}} \phi\left(z \middle| \tilde{B}_{it} x_{it}, \tilde{\Sigma}_{it}\right) dz \tag{A.19}$$

where $J^* = n(E_{it})$. We set the region $a_j,\ j = 1, \cdots, J^*$ as $a_j = (\max(z_j, 0), \infty)$ and $a_k = \left(-\infty, \max(z_j, 0)\right]$ if $k \neq j$ if to obtain the choice probability of j-th alternatives, besides $a_k = (-\infty, 0], k = 1, \cdots, J^*$ if to obtain the external search probability. Let the choice probabilities of alternatives in the experiential set be $p_1, \cdots, p_{J^*}$ respectively and the external search probability be q. Thus, $p_1 + \cdots + p_{J^*} + q = 1$.

Since this model has no information about the alternative choice after the external search, we theoretically cannot forecast which will be chosen after the external search. Note that, as mentioned in Simon (1947), individual search and choice are not a completely random. Consumers tend to choose the alternative they will prefer and can predict what this will be to some extent. We then fill the choice probabilities of the alternatives after the external search from the hierarchical structure. We will now describe the procedure we used to fill the probabilities.

Since we have the preference parameter β_j, even if the alternative is not in the experiential set, we use that parameter. Let the number of alternatives not in the experiential set be J^- and the corresponding extracted parameters be $\breve{B}_{it}$ and $\breve{\Sigma}_{it}$. From these parameters, we obtain the choice probability of these alternatives from the following equation:

$$\Pr\left(y_{itj} = 1 \middle| j \notin E_{it}\right) = \int_{a_1} \cdots \int_{a_{J^-}} \phi\left(z \middle| \breve{B}_{it} x_{it}, \breve{\Sigma}_{it}\right) dz \tag{A.20}$$

where $a_j = (\max(z_j), \infty)$ and $a_k = (-\infty, \max(z_j)]$ if $k \neq j$ if to obtain the choice probability of j-th alternatives. Let the obtained probability be $\breve{p}_j$. Note that the sum $P_s = \breve{p}_1 + \cdots + \breve{p}_J$ will not be 1. Then, divide by the sum P_s, and define $\breve{p}_j / P_s$ as the conditional choice probability of the j-th alternative. We define the choice probability of the j-th alternative not in the experiential set by multiplying the external search probability q. We can then obtain the set of choice probabilities whose sum is 1:

$$\sum_{j \in E_{it}} p_j + q \sum_{j \notin E_{it}} \frac{\breve{p}_j}{P_s} = 1 \qquad \text{(A.21)}$$

The first term on the left hand side is the summary of the choice probabilities of the alternatives in the experiential set. The second term is the summary of the choice probabilities of the alternatives not in the experiential set. In the second term, all choice probabilities are multiplied by the external search probability.

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.

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.

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.

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

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

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.

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

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

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.

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

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

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

Imai, K., and van Dyk, D.A. (2005). A Bayesian Analysis of the Multinomial Probit Model Using Marginal Data Augmentation, *Journal of Econometrics*, 124, 311-334.

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

Nobile, A. (1998). A Hybrid Markov Chain for Bayesian Analysis of the Multinomial Probit Model, *Statistics and Computing*, 8, 229-242.

Oliver, R. L. (2010) *Satisfaction: A Behavioral Perspective on the Consumer 2nd ed.*, New York: M.E. Sharpe.

Petty, R. E., & Cacioppo, J. T. (1986). Communication and persuasion: Central and peripheral routes to attitude change. New York: Springer-Verlag.

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.

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.

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.