ESSAYS ON TIME-VARYING CONSUMER PREFERENCES

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Sungho Park

May 2010

ESSAYS ON TIME-VARYING CONSUMER PREFERENCES

Sungho Park, Ph. D.

Cornell University 2010

Consumer preferences are changing over time.    In this dissertation, we provide three studies regarding changes in consumer preferences and methods of modeling time-varying preferences.

In Chapter 1, we propose a Simulated Maximum Likelihood estimation method for the random coefficient logit model using aggregate data, accounting for heterogeneity and endogeneity.    Our method allows for two sources of randomness in observed market shares – unobserved product characteristics and sampling error. Because of the latter, our method is suitable when sample sizes underlying the shares are finite.    We show that the proposed method provides unbiased and efficient estimates of demand parameters.    We also obtain endogeneity test statistics as a by-product, including the direction of endogeneity bias.    The model can be extended to incorporate Markov regime-switching dynamics in parameters and is open to other extensions based on Maximum Likelihood.    The benefits of the proposed approach are achieved by assuming normality of the unobserved demand attributes, an assumption that imposes constraints on the types of pricing behaviors that are accommodated.    However, we find in simulations that demand estimates are fairly robust to violations of these assumptions.

We propose a structural model of market evolution and apply the proposed model to the South Korean cigarette market data in Chapter 2.    In the South Korean

cigarette market, consumers have shown dramatic changes in their cigarette preferences.   While most consumers smoked high-tar cigarettes ten years ago, now most consumers prefer low-tar cigarettes.   Another interesting trend in this market is the growing popularity of super-slim cigarettes.   Given the strong dynamics in preferences, we raise two critical questions – 1) what are the sources of preference change, and 2) how does the firm (KT&G Corporation, a de-facto monopolist in the market) react to these preference changes.   We answer these questions using a unique structural model of consumer demand and firm behavior.   In the proposed demand model, evolution of consumers' preferences is driven by an exogenous effect and a new product introduction effect.   On the one hand, the increasing preference for low-tar cigarettes can be explained by consumers' growing heath consciousness, an exogenous effect.   Due to stringent government restrictions on promotion and advertising of tobacco products, new product introduction is an important marketing instrument for KT&G.   We hypothesize that a new product carries critical information that subsequently influences consumer preferences.   This is the introduction effect.   We propose an aggregate random coefficient logit model wherein the parameters evolve as a function of the introduction and exogenous effects. This model allows us to separate the two effects and examine their relative significance.   Another key research question we study is how the firm reacts to the preference changes.   To answer this question, we build two supply side models. First, we specify the firm's pricing model which elucidates the influence of the time-varying preferences on the firm's pricing decisions.   Second, we model the firm's decisions regarding new product design and introduction.   This model clarifies the firm's decision process regarding the new product under the time-varying consumer preferences.   This study provides valuable insights into the sources of preference changes, and how firms' decisions shape the fundamentals of the market.   Also, it

sheds light on the role and the value of new products design and introduction.    The proposed model can help a firm develop a new product strategy that will move consumer preferences in a preferred direction.

In many categories consumers display cyclical buying: they repeatedly purchase in the category for several periods, followed by several periods of not buying. One possible explanation for such cyclicality is the joint effect of habit and boredom on repeated purchasing.    In Chapter 3, we propose a Markov regime-switching random coefficient logit model to represent these behaviors as stochastic switching between high and low category purchase tendencies.    The main feature of the proposed model is that it divides the stream of purchase decisions of a consumer into distinct regimes with different parameter values that characterize high versus low purchase tendencies.    In an empirical application of the model to purchases of yogurt-buying households we find that as many as 40.8% display cyclicality between high and low yogurt purchasing tendencies.    We show (via simulation) that alternating between high and low purchase tendencies corresponds with changing levels of consumer inventory in a substitute category.    If one ignores this phenomenon, a correlation between yogurt inventory and the unexplained part (or error term) in utility arises leading to biased estimates.    Predictions from the proposed model track observed yogurt purchases of households over time closely, and the model also fits better than three benchmark models.    Also, we show that cyclicality in buying has a key implication for a firm's price promotion strategies: a price reduction that is offered to a household during its high purchasing tendency period will result in greater increases in sales than one that is offered during its low purchasing period.    This opens up a new dimension for enhancing the effectiveness of promotions - customized timing of price reductions.

BIOGRAPHICAL SKETCH


Sungho Park was born on October 17th, 1976 in Gwangju, South Korea.   In his early years, he enjoyed learning and studying different languages.   As such, he chose to attend Myung Duk Foreign Language High School, a magnet school dedicated to the study of various foreign languages, where he studied Japanese and Chinese in addition to the mandatory curricula in English.   Sungho then matriculated to Seoul National University, where he majored in Linguistics.   During his undergraduate studies, his intellectual curiosity expanded to various fields in the humanities, natural sciences, and social sciences.   In his senior year, after returning to the university from a mandatory 26-month military service, he happened to take courses in statistics and management science.   Those courses piqued his interest in quantitative methods and their application to managerial problems, and proved to be the catalyst for Sungho to change directions in his academic pursuits.   Upon graduating from Seoul National University, he decided to matriculate to the Master of Science program in management engineering at the Korea Advanced Institute of Science and Technology (KAIST).   At KAIST, he became a member of the Business Forecasting Lab, which performs research on statistical forecasting methods in business and economics under the advisement of Professor Duk Bin Jun.   There, his research focused on various statistical techniques to handle marketing problems, and he also had the opportunity to apply those techniques to real world marketing problems in projects with several leading South Korean companies, including Samsung and Korea Telecom.   Sungho's experiences in marketing research while at KAIST strongly motivated him to pursue further research in academia, as a result of which, he decided to join the PhD program in quantitative marketing at the Johnson Graduate School of Management at Cornell University.   While at Johnson, Sungho

has very much enjoyed the intimate environment of the marketing group which, in his opinion, has fostered more productive interaction with faculty.    Under the advisement of Professor Sachin Gupta, he performed research on the dynamic aspects of consumer preferences and related marketing issues.

To my parents, Jong-Shick Park and Keum-Sun Kim,

and to my wife, Rebecca In-Young Chung

ACKNOWLEDGMENTS


Professor Sachin Gupta, thank you very much for your support and encouragement throughout my PhD program.    I gained an enormous amount of knowledge from you in every stage of research.    You are my role model in both research and teaching.    I deeply appreciate the valuable advice that you have provided me for the past 5 years.    I am, and will continue to be, very proud of being a member of your academic genealogy.

Professors Edward McLaughlin, Young-Hoon Park, and Vishal Narayan, thank you for being my committee members and providing thoughtful advice on my research. Also, I deeply appreciated your support in my job search process.

Professor Vithala Rao, thank you for careful advice on my research.    Your passion in teaching and research motivated me throughout my PhD program.

Thanks as well to the faculty members of marketing group at the Johnson Graduate School of Management: Professors Alice Isen, Vrinda Kadiyali, J. Edward Russo, Doug Stayman, and Thomas Manoj.

Finally, I thank my wife Rebecca, my parents and my sister Sun-Young for their continuous support and encouragement.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLE

CHAPTER 1

A SIMULATED MAXIMUM LIKELIHOOD ESTIMATOR FOR THE RANDOM

COEFFICIENT LOGIT MODEL USING AGGREGATE DATA

**1.1 Introduction**

In the estimation of market demand, heterogeneity across consumers and the endogeneity of marketing activities, especially price, are major concerns of marketing researchers. It has been reported that ignoring heterogeneity and/or endogeneity causes a bias in demand estimates (Berry 1994; Keane 1997; Besanko, Gupta, and Jain 1998; Villas-Boas and Winer 1999; Chintagunta 2001; Chintagunta, Dubé and Goh 2005). Recognizing their importance, researchers have tackled both issues in aggregate models and disaggregate models.[1]

In disaggregate models, several estimation methods have been suggested and compared (Villas-Boas and Winer 1999; Petrin and Train 2004; Draganska, and Jain 2002; Yang, Chen, and Allenby 2003; Goolsbee and Petrin 2004; Chintagunta, Dubé, and Goh 2005). For aggregate models, Berry, Levinsohn, and Pakes (1995) developed a method (henceforth, BLP method) that provides consistent estimates under heterogeneity and endogeneity. The BLP method has been applied successfully in numerous studies (Sudhir 2001a; Petrin and Train 2002) and has become the most widely used approach for analyzing differentiated product markets.

A distinguishing feature of the BLP model is that it assumes that the observed market shares of alternatives have *no* sampling error. Randomness in shares in the BLP model is assumed to come *only* from unmeasured product characteristics. If the data being modeled contain more than minimal sampling error, the BLP estimator is

---

[1] In this paper, we refer to a model that uses aggregate- or market-level data as an aggregate model, and a model that uses household- or individual-level data as a disaggregate model. Within this definition, the aggregate model can specify utility at the household- or individual-level.

not consistent and asymptotically normal (Petrin and Train 2004; Berry, Linton, and Pakes 2004). The BLP model was originally applied to automobile shares for the entire United States market in which the number of households was of the order of 100 million; hence, sampling error was negligible. In many subsequent applications of the model to weekly supermarket point-of-sale data (two examples are Chintagunta 2002 and Sriram et al. 2007), the underlying sample of shopper households is quite large, thus satisfying the no-sampling-error assumption of the BLP model.

In marketing one can identify a number of situations wherein the assumption of negligible or no sampling error in observed brand shares may not be tenable. These occur when the sample of shoppers underlying the observed shares is relatively small. Examples of such situations include the following:

a) Sales data from smaller retail stores;

b) Sales data for infrequently purchased categories;

c) Sales data at the stock keeping unit (SKU) level, which by definition have smaller sales than brands or brand-sizes;

d) Shares computed using aggregated household panel data. This may be necessary if the household-level data cannot be used due to, for instance, privacy concerns; and

e) Household panel data are aggregated to estimate brand shares because point-of-sale data are unavailable (e.g. Walmart does not provide point-of-sale data to ACNielsen or Information Resources Inc.).

In all these situations the assumptions of the BLP model may not met. We propose in this paper a Simulated Maximum Likelihood (SML) method to estimate an aggregate random coefficient logit model that considers endogeneity as well as heterogeneity. Our method is suitable for share data that are observed with sampling error. Thus, we assume that there are two sources of randomness in the model – unmeasured product characteristics, and sampling error. Our proposed method is

motivated by the control function approach which was originally suggested for the disaggregate model by Villas-Boas and Winer (1999) and later extended to the aggregate model by Petrin and Train (2004). Villas-Boas and Winer (1999) developed their model for individual data and did not allow for unobserved individual heterogeneity. By contrast, our approach is for aggregate data, and we model unobserved heterogeneity using a random coefficients framework. In relation to Petrin and Train (2004), our model makes different assumptions on the distribution of the unobservables. We elaborate upon this distinction when we discuss the model in a subsequent section.

Using simulated data we demonstrate that the proposed estimator provides unbiased and efficient estimates of demand parameters. The estimation procedure is straightforward to understand and implement. Furthermore, the proposed method can readily incorporate other methods based on Maximum Likelihood Estimation (MLE). For example, we can incorporate Markov regime-switching models (or hidden Markov models) into our framework. By doing so, we can investigate parameter dynamics in choice models using aggregate data when both heterogeneity and endogeneity are present. A further benefit of our proposed model is that an endogeneity test statistic results as a by-product. A test for endogeneity based on the Wald statistic or the Likelihood Ratio statistic can then be easily performed.

In the proposed approach, we impose structure on the distribution of unmeasured product characteristics by making a normality assumption. We find that this leads to efficient estimates of heterogeneity parameters which are of great practical interest in marketing applications such as segmentation and targeting. The distributional assumptions we make impose restrictions on the types of pricing behaviors that are accommodated (we elaborate upon this later), although we find in simulations that demand estimates are fairly robust to violations of these assumptions.

3

Chintagunta, Dubé, and Goh (2005) showed that even when there is no price endogeneity, researchers have to pay attention to the presence of unmeasured product characteristics that affect consumer utility. Unmeasured product characteristics may include, for example, the impact of unobserved promotional activity, coupon availability, shelf space, national advertising, unquantifiable factors and systematic shocks to demand. If omitted from the model, the unmeasured product characteristics generate overstated variances in the estimated distribution of heterogeneity in household brand preferences and price sensitivities. An additional contribution of our paper is to expand upon this important finding of Chintagunta, Dubé, and Goh (2005). We show that problems due to the omission of the unmeasured product characteristics are more complex, and have additional facets which have not been reported in the literature. In particular, the omission can cause upward or downward biases in mean and/or heterogeneity parameters.

The remainder of the paper is organized as follows. In the next section, we review related literature. Following that, we present the model and explain our estimation method. We then evaluate the performance of the proposed method in simulation studies. In the next section, we apply the proposed method to scanner panel data and compare results with those from extant methods. We conclude in the last section.

## 1.2 Literature Review

We focus on literature that tackles endogeneity as well as heterogeneity in choice models.[2] In disaggregate models the available methods can be classified into three categories: 1) full-information maximum likelihood approaches (Sudhir 2001b; Draganska and Jain 2002; Yang, Chen, and Allenby 2003; Villas-Boas and Zhao

---

[2] We do not include here papers that tackle a related form of endogeneity in which marketing variables are set as a function of consumer responsiveness (e.g. Manchanda et al. 2004) or cross-sectional sales differences (e.g. Bronnenberg and Mahajan 2001).

2005), 2) control function approaches (Villas-Boas and Winer 1999, Petrin and Train 2004), and 3) fixed-effect approaches (Goolsbee and Petrin 2004; Chintagunta, Dubé, and Goh 2005). In the full-information maximum likelihood approach, prices are modeled as the equilibrium outcome of a game between firms. By explicitly modeling price, one can integrate out unmeasured product characteristics and derive the unconditional joint likelihood of prices and choices. The control function approach is based on the concepts of Heckman (1978) and Hausman (1978), or it can be viewed as a reduced-form approximation of the equilibrium model. This approach requires two steps. First, the endogenous variable is regressed on instrumental variables. Second, the residual from the first step regression, or a function of the residual, is entered as an additional explanatory variable in utility to control for unmeasured product characteristics. In the fixed-effect approach, the first step is to capture the endogeneity by product- and/or market-specific fixed-effects and then, in the second stage, a standard instrumental variables method is applied to these fixed-effects. Chintagunta, Dubé, and Goh (2005) directly estimate the product- and market-specific fixed effects using MLE. Goolsbee and Petrin (2001) adopt the numerical inversion method (or contraction mapping) suggested by Berry (1994) to get the fixed effects.

There is a growing stream of work in marketing and economics that uses aggregate data to estimate choice models. One obvious reason for this trend is easier availability of aggregate data. A widely used approach for dealing with endogeneity as well as heterogeneity is the fixed-effect approach which was first developed for the aggregate model by BLP (1995) and later applied to a disaggregate model. Unlike the disaggregate model, in an aggregate model we cannot directly estimate the fixed-

effects due to lack of degrees of freedom.[3] The BLP method circumvents the direct estimation of fixed-effects by using a numerical inversion method instead.

A weakness of the BLP method is its inability to recover heterogeneity parameters precisely when only aggregate data are used (Petrin 2002; Albuquerque and Bronnenberg 2006). Petrin (2002) proposed a technique to augment aggregate data with information relating consumer demographics to the characteristics of the products these consumers purchase. Similarly, Albuquerque and Bronnenberg (2006) supplement aggregate data with summaries of household switching behavior. An important strength of the BLP method is that it makes few assumptions about the distribution of unobserved product characteristics. The only assumption is that the unobserved characteristics are mean independent of the instrumental variables. As a result, the BLP method does not impose restrictions on the form of pricing behavior.

A number of recent papers perform Bayesian analysis of the random coefficient logit model using aggregate data. Musalem, Bradlow, and Raju (2007) consider two alternative scenarios that generate the observed aggregate data – one in which there are independent cross-sections of consumers in each period, and the second in which there is a panel of consumers. They note computational limitations of their approach when the number of individual consumers underlying the aggregate data is larger than about 500. Their second scenario is similar to the one in Chen and Yang (2006) who propose a data augmentation approach to capture household heterogeneity. However, Chen and Yang do not consider unmeasured product characteristics or related price endogeneity issues, both of which are crucial to our research goals.

---

[3] Say we consider *J* inside alternatives and *T* markets. Since the *J+1*th alternative (outside good) is normalized, degrees of freedom in aggregate data are *J×T,* which is the number of fixed-effects to be estimated.

Recently, Jiang et al. (2007) propose a Bayesian analysis of the aggregate random coefficient logit model based on distributional assumptions about the unmeasured product characteristics. Similar to the BLP method, this approach is suitable when there is no sampling error in observed shares. Unlike our proposed SML method, model estimation in their approach requires inverting shares via the BLP contraction mapping as well as relatively complicated Markov-Chain Monte Carlo sampling. Like us, Jiang et al. demonstrate via numerical simulation that under misspecification of the distribution of the unmeasured product characteristic, their method continues to produce good results. In general, however, the properties of Bayesian estimators under model misspecification are not well established. By contrast, MLE is a strongly consistent estimator that minimizes the Kullback-Leibler Information Criterion (KLIC) (White 1982). That is, the proposed method provides estimates which are closest in KLIC to the true parameters in vector space defined by the normal approximation. In this respect we believe our method is more robust than Bayesian approaches.

**1.3 Model and estimation procedure**

**1.3.1 Model**

Our interest is in consistent and efficient estimation of the random coefficient brand choice model under assumptions of heterogeneity across consumers and endogeneity of marketing activities. We assume that consumers either choose a single unit of the brand that gives them the highest utility in the category or choose not to purchase in the category on a given shopping trip. In this paper, we focus on purchase incidence and brand choice behaviors only. In each week $t=1,...,T$, the utility of brand $j=1,...,J$ for consumer $h=1,...,H$ is given by the following expression:

$$u_{hjt} = x'_{jt}\beta_h + \xi_{jt} + \varepsilon_{hjt}, \tag{1}$$

$$u_{h(J+1)t} = \varepsilon_{h(J+1)t}, \text{ if no purchase,}$$

where $x_{jt}$ is a *k*-dimensional vector of observed marketing mix variables and intrinsic brand values (brand intercepts), $\beta_h$ is a *k*-dimensional vector of individual specific tastes for characteristics and marketing mix responsiveness, $\xi_{jt}$ is unmeasured product characteristics that are unobserved by the researchers but considered by consumers in their purchase decisions and by marketers in their decision making, and $\varepsilon_{hjt}$ is an i.i.d. random shock with a Type-I Extreme Value distribution. Consumer preferences are heterogeneous and to capture this, we model the taste vector $\beta_h$ as a random draw from a multivariate normal distribution $N(\overline{\beta}, \Omega)$:

$$\beta_h = \overline{\beta} + \Omega^{1/2}\eta_h, \quad \eta_h \sim N(0, I_k), \tag{2}$$

where $\Omega^{1/2}$ is the lower-triangular Cholesky factor of $\Omega$, $\overline{\beta}$ is the mean parameter of the distribution of heterogeneity, and $\Omega^{1/2}\eta_h$ is individual-specific deviation from the mean.

We can raise two issues related to $\xi_{jt}$. The first issue is the endogeneity problem. If marketers make their decisions based on the values of $\xi_{jt}$, marketing mix variables in $x_{jt}$ would be correlated with $\xi_{jt}$. In particular, empirical research has repeatedly reported the correlation between price and $\xi_{jt}$ (or price endogeneity) in disaggregate as well as in aggregate data. Due to this correlation, $\xi_{jt}$ is not necessarily mean zero given $x_{jt}$ and thus, we cannot treat it as another error component and integrate it out of the demand function. Second, regardless of the correlation with $x_{jt}$, ignoring $\xi_{jt}$ would force the model to absorb these effects in the i.i.d. random shock and/or the explained part of the utility $x'_{jt}\beta_h$. As a result, one could get biased estimates of model parameters.

Following Heckman (1978) and Hausman (1978), we explicitly introduce the endogeneity issue into the random coefficient logit model by the following specification:

$$x_{jt} = (I_k \otimes z'_{jt})\gamma_j + v_{jt}, \quad v_{jt} \sim i.i.d. \, N(0, \Sigma_{v_j}), \tag{3}$$

$$\xi_{jt} \sim i.i.d. \ N(0, \sigma^2_{\xi_j}), \tag{4}$$

$$Cov(v_{jt}, \xi_{jt}) = \lambda_j, \tag{5}$$

$$Cov(z_{jt}, \xi_{jt}) = 0 \ \forall t, \tag{6}$$

where $z_{jt}$ is an *L*-dimensional vector of instrumental variables uncorrelated with $\xi_{jt}$ but correlated with $x_{jt}$. $z_{jt}$ includes exogenous variables in $x_{jt}$. The distributional assumptions in (3) and (4) allow us to directly apply SML estimation, as described next. Without loss of generality, we apply the Cholesky decomposition of the covariance matrix of $[v'_{jt} \ \xi'_{jt}]'$ in order to rewrite it as a function of two independent shocks:

$$\begin{bmatrix} v_{jt} \\ \xi_{jt} \end{bmatrix} = \begin{bmatrix} b_{11,j} & 0 \\ b_{21,j} & b_{22,j} \end{bmatrix} \begin{bmatrix} \omega_{1,jt} \\ \omega_{2,jt} \end{bmatrix}, \quad \begin{bmatrix} \omega_{1,jt} \\ \omega_{2,jt} \end{bmatrix} \sim i.i.d. N\left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} I_k & 0 \\ 0 & 1 \end{pmatrix} \right), \tag{7}$$

$$Cov(\omega_{1,it}, \omega_{1,jt}) = 0 \quad for \ i \neq j,$$

$$Cov(\omega_{2,it}, \omega_{2,jt}) = 0 \quad for \ i \neq j,$$

where $b_{11,j} = \Sigma_{v_j}^{1/2}$. Using (7), we can rewrite (1) and (3) as follows:

$$u_{hjt} = x'_{jt} \beta_h + b_{21,j} \omega_{1,jt} + b_{22,j} \omega_{2,jt} + \varepsilon_{hjt}, \tag{8}$$

$$x_{jt} = (I_k \otimes z'_{jt}) \gamma_j + b_{11,j} \omega_{1,jt}, \tag{9}$$

Solving (9) for $\omega_{1,jt}$ and substituting this in (8) results in the following transformation of (1):

$$u_{hjt} = x'_{jt} \beta_h + \theta'_j (x_{jt} - (I_k \otimes z'_{jt}) \gamma_j) + b_{22,j} \omega_{2,jt} + \varepsilon_{hjt}, \tag{10}$$

where $\theta'_j = b_{21,j} b_{11,j}^{-1}$. Recall that $\varepsilon_{hjt}$ is an i.i.d. random shock with a Type-I Extreme Value distribution and $\omega_{2,jt}$ is an i.i.d. random shock with a standard normal distribution. More importantly, $\omega_{2,jt}$ is uncorrelated with any other term in (10); we refer to this term as the exogenous unmeasured product characteristic or EUPC. In our approach we treat $\omega_{2,jt}$ as an additional error component and integrate it out of the

demand function. Also note that $(x_{jt} - (I_k \otimes z'_{jt})\gamma_j)$ is the residual from the regression in (3) and works as a bias correction term.

Our model is closely related to the "control function" method of Petrin and Train (2004). This method is based on a decomposition of the unmeasured product characteristic term as follows: $\xi_{jt} = f_j(v_t) + \varpi_{jt}$ where $v_t = [v'_{1t} \cdots v'_{Jt}]'$. The "control function" $f_j(v_t)$ is a function of the residuals $v_t$ obtained from the first stage regression in (3); this term controls for endogeneity. The new error components $\varpi_{jt}$ are similar to our EUPC $\omega_{2,jt}$. While the proposed method assumes joint-normality of $v_{jt}$ and $\xi_{jt}$, the control function method requires specification of the functional form of $f_j(\cdot)$ and the distribution of $\varpi_{jt}$. Note that we can get (10) from Petrin and Train's model by letting $f_j(v_t) = \theta'_j v_{jt}$ and $\varpi_{jt} = b_{22,j}\omega_{2,jt}$ along with the assumption of normality. In an empirical application of their model to the original automobile data of BLP (1995), Petrin and Train (2004) specified $\varpi_{jt}$ to be Normal and hence this term was not separately identifiable from the normal random deviate in the constant term in utility. As a result the term did not need to be handled separately. However, in our model specification of brand choice which is popular in marketing, we need to separately integrate out the EUPC, as we explain in the next subsection.

### 1.3.2 Estimation procedures and endogeneity tests

From (2) and (10) we can derive the logit probability that consumer $h$ chooses alternative $j$:

$$P_{hjt} = \frac{\exp(x'_{jt}\overline{\beta} + \theta'_j(x_{jt} - (I_k \otimes z'_{jt})\gamma_j) + b_{22,j}\omega_{2,jt} + x'_{jt}\Omega^{1/2}\eta_h)}{1 + \sum_{i=1}^{J}\exp(x'_{it}\overline{\beta} + \theta'_i(x_{it} - (I_k \otimes z'_{it})\gamma_i) + b_{22,i}\omega_{2,it} + x'_{it}\Omega^{1/2}\eta_h)} \tag{11}$$

(11) has the usual random coefficient logit form except a bias correction term $(x_{jt} - (I_k \otimes z'_{jt})\gamma_j)$ and time- and alternative-specific shocks $\omega_{2,t} = [\omega_{2,1t} \cdots \omega_{2,Jt}]$. Now we will describe a way to handle these shocks in the estimation. For expositional convenience, we first assume that the bias correction term, $(x_{jt} - (I_k \otimes z'_{jt})\gamma_j)$, is given.

10

Conditional on $\omega_{2,t}$, we can write the likelihood of the observed aggregate data in week $t$:[4] (Note that the assumption of a multinomial sampling process is made here, resulting in sampling error.)

$$L_{1,t}(\omega_{2,t}) = \left(\frac{H!}{n_{0t}! \cdots n_{J+1t}!}\right) \prod_{j=1}^{J+1} \left(\int P_{hjt}(\omega_{2,jt}, \eta_h) \phi(\eta) \, d\eta\right)^{n_{jt}}, \qquad (12)$$

where $n_{jt}$ is the count of purchase trips for brand $j$ in week $t$ and $\phi(\cdot)$ is the standard normal density function. Since $\omega_{2,t}$ are unknown, we again integrate them out:

$$L_{1,t} = \int L_{1,t}(\omega_{2,t}) \phi(\omega_2) \, d\omega_2, \qquad (13)$$

and the likelihood function for the sample of $T$ weeks is

$$L_1 = \prod_{t=1}^{T} L_{1,t}. \qquad (14)$$

For the computation of (12) and (13), we can use Monte Carlo simulation methods or SML (see Keane 1993).

In the implementation of SML, evaluation of the likelihood may encounter computational difficulties when $n_{jt}$ is "large", because $\left(\int P_{hjt}(\omega_{2,jt}, \eta_h) \phi(\eta) \, d\eta\right)^{n_{jt}}$ in (12) reaches machine zero fairly quickly. Although we were able to apply SML without this computational problem in our empirical application to aggregate data based on paper towel purchases of 880 households (discussed in a subsequent section), the problem is inescapable when $H$ is large. The incidence of this computational problem depends on the size of $H$ (or $n_{jt}$), distributions of choice probabilities, and the definition of machine zero on the particular computer and language used for estimation. Our approach to circumvent this problem, when it occurs, is to represent $H$ consumers with a sample of tractable size, $R$. We use the observed sales in each time period $t$ to compute shares of each of the $J+1$ products. We then draw a

---

[4] For the rigorous derivation of this likelihood function, see Bodapati and Gupta (2004).

11

multinomial sample of size $R < H$ from these shares. This new, smaller sample is used to compute the likelihood function and obtain SML estimates[5].

A natural question arises regarding the new sample size: What is the optimal $R$? As $R$ increases, we may expect efficiency gain. However, potential numerical inaccuracy also increases due to the increased exponent in (12). By trial and error with many different values of $R$ ranging from 50 to 500 in simulation experiments, we determined that we get highly satisfactory results with $R=100$ but also note that the results do not change much with $R$. We use $R=100$ in all our simulation studies in Section 4. We also applied the proposed method to many datasets generated from different values of $H$ ranging from 1,000 to 100,000 and get satisfactory results in all cases. As $H$ becomes larger, estimates are distributed closer to the true values but only marginally.

So far, we have assumed that the bias correction term, $(x_{jt} - (I_k \otimes \zeta'_{jt})\gamma_j)$, is given. However, $\gamma_j$ needs to be estimated by maximizing the following likelihood function derived from (9):

$$L_{2,j} = (2\pi)^{-k/2} \left| \Sigma_{v_j} \right|^{-1/2} \exp\left( -0.5(x_{jt} - (I_k \otimes \zeta'_{jt})\gamma_j)' \Sigma_{v_j}^{-1}(x_{jt} - (I_k \otimes \zeta'_{jt})\gamma_j) \right). \quad (15)$$

A joint estimation of the model can be performed by maximizing the following log likelihood function:

$$\ln L = \ln L_1 + \ln L_{2,1} + \cdots + \ln L_{2,J}. \quad (16)$$

In the above setting, a test for endogeneity is easy to perform. Note that $b_{21,j}$ captures the correlation between $\xi_{jt}$ and $x_{jt}$. If there is no correlation between $\xi_{jt}$ and $x_{jt}$, then $b_{21,j} = 0$ and $\theta'_j = b_{21,j}b_{11,j}^{-1} = 0$. Since our estimation procedure is based on SML, we can apply the standard hypothesis testing framework of MLE.[6] The null hypothesis

---

[5] Statistical properties of this estimator are provided in Appendix 1.
[6] If the number of draws in the simulation rises faster than the sample size, SML is consistent, asymptotically normal and efficient, and equivalent to ML (Train 2003 p.259), justifying our application of the standard hypothesis testing framework of ML.

(i.e. no endogeneity) is $H_0 : \theta = [\theta'_1 \quad \cdots \quad \theta'_J]' = 0$. The likelihood ratio (LR) test

statistic and the Wald statistic can be derived as follows:

$$Wald = \hat{\theta}' Cov(\hat{\theta})^{-1} \hat{\theta} \sim \chi^2(Q),$$

$$LR = -2(\ln L_R - \ln L_{UR}) \sim \chi^2(Q),$$

where $\ln L_R$ and $\ln L_{UR}$ are the log likelihood value with- and without-restriction,

respectively, and $Q$ is the dimension of $\theta$. More simply, we can obtain the

significance of $\theta$ directly from the estimation result. We can regard this test as an

extension of a regression-based Hausman test (Hausman 1978, 1983; also see

Wooldridge 2001, p.118) or Wu test (Wu 1973) to a random coefficient logit model.

### 1.3.3 Implications of the Assumption of Joint Normality of $\xi$ and $\nu$

The assumption of joint normality of unmeasured product characteristics $\xi$

and price residuals $\nu$ in the proposed method (equations (3) and (4)), while standard

from a statistical perspective, has strong economic implications. In particular, when

the endogenous explanatory variables are prices, this assumption is inconsistent with

many forms of pricing behavior.

To explain this issue we begin with an example where the normality

assumption *is consistent* with pricing. Rewriting the utility function in (1) by

redefining $x_{jt}$ to contain only non-price observed attributes, we have

$$u_{hjt} = \alpha_h p_{jt} + x'_{jt} \beta_h + \xi_{jt} + \varepsilon_{hjt}$$

where $p_{jt}$ is the price of product $j$ at $t$. Let the marginal cost of product $j$ be linear in

the observed and unobserved non-price attributes plus an error representing

unobserved cost shocks:

$$MC_{jt} = x'_{jt} \gamma + \lambda \xi_{jt} + \zeta_{jt}$$

Suppose that each product is priced at marginal cost, as in perfect competition. Then

the price equation becomes

$$p_{jt} = x'_{jt} \gamma + \lambda \xi_{jt} + \zeta_{jt} = x'_{jt} \gamma + \nu_{jt}$$

In this situation the assumption of a normal distribution for unmeasured product

characteristics $\xi$ and a normal distribution for the marginal cost shocks $\zeta$ implies a

joint normal for the error in the pricing equation $v$ and the unmeasured product

characteristics $\xi$. This is also the case when prices are equal to marginal cost plus a

fixed markup. Other forms of pricing do not yield this result. For example, under two

prominent theories of pricing -- monopoly pricing and Nash pricing in a differentiated

products oligopoly -- prices are some markup over marginal cost, where the markup

depends on elasticities of demand at the prices. The pricing equation is

$$p_{jt} = x'_{jt}\gamma + MK_{jt}(p,x,\xi) + \lambda\xi_{jt} + \zeta_{jt}$$

where *MK* denotes the profit maximizing markup. There is no way that this pricing

equation can be neatly expressed in the form $p_{jt} = x'_{jt}\tau + v_{jt}$ with a normal distribution

for $v$. The distribution of $v$ is defined implicitly by the solution to the pricing

equation which has prices on both sides. It is not a simple task to derive a distribution

for $v$ from assumed distributions for $\xi$ and $\zeta$. Even if we derive the distribution, it

is not guaranteed to be normal. Furthermore, the distribution of $v$ will not be

independent of $x$.

The foregoing discussion shows that the normality assumption is not

inconsequential. Two factors mitigate the severity of the consequences in practice.

First, the cost plus fixed markup model of pricing is widely practiced; Shim and Sudit

(1995) report that it is used by over 80% of managers at manufacturing firms. Second,

we find in our simulation studies that demand estimates from the proposed model are

quite robust to violations of distributional assumptions.

**1.4 Simulation study**

We conduct simulation experiments with the following goals: a) to assess the

performance of the proposed SML method; b) to assess robustness of the proposed

method to key distributional assumptions about the unmeasured product characteristics,

and c) to assess the implications of omitting the exogenous unmeasured product characteristics (EUPC) from the estimating model.  To achieve these goals we consider five cases.  In each of the five cases 100 datasets are generated as replicates, and models are estimated on each dataset to obtain the empirical sampling distribution of the parameter estimates.  Data are generated from a sample of $H$=100,000 households.[7] We implement SML with $R$=100 as described in Section 1.3.2.

## 1.4.1 Case 1: Simple heterogeneity

We generate data by (1)-(6).  The specific data generating process (DGP) and the parameter values we assign are summarized below:

$$u_{hjt} = x'_{jt}\beta_h + \xi_{jt} + \varepsilon_{hjt} , \quad u_{h(J+1)t} = \varepsilon_{h(J+1)t} , \text{ if no purchase, } \varepsilon_{hjt} \sim i.i.d. \ EVI , \quad (17)$$

$$x_{jt} = [c_{1jt} \quad c_{2jt} \quad x_{1jt} \quad x_{2jt}]' , \quad (18)$$

$$c_{1jt} = I(j=1), \quad c_{2jt} = I(j=2), \quad (19)$$

$$x_{1jt} = z_{1jt} + z_{2jt} + \omega_{1jt} , \quad x_{2jt} = I(\upsilon_{jt} > 0.7), \quad \upsilon_{jt} \sim Unif(0,1), \quad (20)$$

$$\xi_{jt} = \omega_{1jt} + \omega_{2jt} , \quad (21)$$

$$\omega_{1jt} , \ \omega_{2jt} , \ z_{1jt} , \ z_{2jt} \sim i.i.d \ N(0, 0.5), \quad (22)$$

$$\beta_h = \begin{bmatrix} \overline{\beta}_1 \\ \overline{\beta}_2 \\ \overline{\beta}_3 \\ \overline{\beta}_4 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_{33} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \eta_h = \begin{bmatrix} 0.2 \\ 0.5 \\ -1 \\ 1 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \eta_h , \ \eta_h \sim N(0, I_4) , \quad (23)$$

where $I(\cdot)$ denotes an indicator function and $I_k$ denotes a $k$-dimensional identity matrix.  Note that $x_{1jt}$ is correlated with $\xi_{jt}$ through $\omega_{1jt}$ and this results in the endogeneity problem.  We assume that individuals have heterogeneous tastes with respect to $x_{1jt}$ only, $J$=2, $T$=50 or 100, and $H$=100,000.  Even though utility is defined and generated at the disaggregate-level, we use only aggregate data for estimation.

---

[7] All five simulations were also conducted with 1,000 households and the substantive findings were identical to those reported here for 100,000 households. Results are available from the authors on request.

We have eight instrumental variables ($c_{1jt}$, $c_{1jt}z_{1jt}$, $c_{1jt}z_{2jt}$, $c_{1jt}x_{2jt}$, $c_{2jt}$, $c_{2jt}z_{1jt}$, $c_{2jt}z_{2jt}$, and $c_{2jt}x_{2jt}$) and five parameters to estimate ($\overline{\beta}_1$, $\overline{\beta}_2$, $\overline{\beta}_3$, $\overline{\beta}_4$, and $\sigma_{33}$). Hence the model is over-identified.

The first two moments of the empirical sampling distributions of the parameter estimates are summarized in Table 1.1. The proposed method works well in the recovery of mean and heterogeneity parameters. Even when the sample is as small as $T=50$, the estimates of mean and heterogeneity parameters are distributed close to the true values and we can conclude that the method provides unbiased estimates.

An additional feature of the proposed method is a simple test of endogeneity. By checking the significance of $\theta\,(=\theta_1=\theta_2)$, we can formally test whether endogeneity is present. Note that in the present DGP, $x_{1jt}$ is correlated with $\xi_{jt}$ (i.e., endogeneity is present). In Table 1.1 we observe that the empirical distribution of $\theta$ is tightly distributed around the true value 1 for both $T=50$ and $T=100$ leading to the correct conclusion that $\theta$ is significantly different from zero and that endogeneity is present.

**Table 1.1: Results of the Simulation Study Case 1 – Simple Heterogeneity (based on 100 replications)**

| Parameters | True Values | T=50 | | T=100 | |
|---|---|---|---|---|---|
| | | Mean | SE | Mean | SE |
| $\overline{\beta}_1$ | 0.200 | 0.212 | 0.136 | 0.206 | 0.100 |
| $\overline{\beta}_2$ | 0.500 | 0.496 | 0.157 | 0.489 | 0.105 |
| $\overline{\beta}_3$ | -1.000 | -1.007 | 0.163 | -0.977 | 0.147 |
| $\overline{\beta}_4$ | 1.000 | 0.990 | 0.205 | 0.980 | 0.158 |
| $\theta_1 = \theta_2$ | 1.000 | 1.060 | 0.185 | 0.980 | 0.117 |
| $\sigma_{33}$ | 1.000 | 1.005 | 0.194 | 1.009 | 0.152 |
| $SD(\omega_{2jt})$ | 0.707 | 0.718 | 0.087 | 0.725 | 0.065 |

### 1.4.2 Case 2: Full heterogeneity

Here we generate data using a more complete heterogeneity distribution. The specific DGP and the parameter values we assign are the same as in Case 1 except for (23). The modified heterogeneity distributions are expressed as follows:

$$
\beta_h = \begin{bmatrix} \bar{\beta}_1 \\ \bar{\beta}_2 \\ \bar{\beta}_3 \\ \bar{\beta}_4 \end{bmatrix} + \begin{bmatrix} \sigma_{11} & 0 & 0 & 0 \\ 0 & \sigma_{22} & 0 & 0 \\ 0 & 0 & \sigma_{33} & 0 \\ 0 & 0 & 0 & \sigma_{44} \end{bmatrix} \eta_h = \begin{bmatrix} 0.2 \\ 0.5 \\ -1 \\ 1 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \eta_h,
$$

$$
\eta_h \sim N(0, I_4). \tag{23a}
$$

Once again, $J=2$, $T=50$ and 100, and $H=100,000$. Note that consumers have heterogeneous tastes with respect to all $x_{jt}$. We have eight instruments ($c_{1jt}$, $c_{1jt}z_{1jt}$, $c_{1jt}z_{2jt}$, $c_{1jt}x_{2jt}$, $c_{2jt}$, $c_{2jt}z_{1jt}$, $c_{2jt}z_{2jt}$, and $c_{2jt}x_{2jt}$) and eight parameters to estimate ($\bar{\beta}_1$, $\bar{\beta}_2$, $\bar{\beta}_3$, $\bar{\beta}_4$, $\sigma_{11}$, $\sigma_{22}$, $\sigma_{33}$, and $\sigma_{44}$). Thus, the model is exactly identified.

The first two moments of the empirical distributions of the parameter estimates are summarized in Table 1.2. The estimates of mean parameters (i.e. $\bar{\beta}_1$, $\bar{\beta}_2$, $\bar{\beta}_3$, and $\bar{\beta}_4$) and heterogeneity parameters (i.e., $\sigma_{11}$, $\sigma_{22}$, $\sigma_{33}$, and $\sigma_{44}$) are close to their true values indicating that the method provides unbiased estimates. However, dispersions of the distribution are larger than those in Case 1. This is due to the additional complexity in heterogeneity distributions of the current DGP. Note also that the empirical distribution of $\theta$ is tightly distributed around the true value 1 for both $T=50$ and $T=100$, supporting the presence of endogeneity in the data, which is the case in our DGP.

**Table 1.2: Results of the Simulation Study Case 2 – Full Heterogeneity (based on 100 replications)**

| Parameters | True Values | T=50 | | T=100 | |
|---|---|---|---|---|---|
| | | Mean | SE | Mean | SE |
| $\bar{\beta}_1$ | 0.200 | 0.187 | 0.219 | 0.227 | 0.137 |
| $\bar{\beta}_2$ | 0.500 | 0.518 | 0.174 | 0.531 | 0.122 |
| $\bar{\beta}_3$ | -1.000 | -1.010 | 0.231 | -0.983 | 0.171 |
| $\bar{\beta}_4$ | 1.000 | 1.019 | 0.336 | 0.966 | 0.215 |
| $\theta_1 = \theta_2$ | 1.000 | 1.026 | 0.252 | 0.991 | 0.175 |
| $\sigma_{11}$ | 1.000 | 1.005 | 0.637 | 0.903 | 0.481 |
| $\sigma_{22}$ | 1.000 | 0.916 | 0.628 | 0.875 | 0.495 |
| $\sigma_{33}$ | 1.000 | 1.033 | 0.266 | 0.993 | 0.190 |
| $\sigma_{44}$ | 1.000 | 1.050 | 0.654 | 0.970 | 0.471 |
| $SD(\omega_{2jt})$ | 0.707 | 0.700 | 0.152 | 0.689 | 0.102 |

### 1.4.3 Case 3: Misspecification due to Autocorrelation in $\omega_{2jt}$

In the DGP we assume autocorrelation in the distribution of the unmeasured product characteristic. All other details of the DGP remain the same as in Case 1 except (22) which is changed as follows:

$$\omega_{2jt} = \phi\,\omega_{2jt-1} + \varsigma_{jt} = 0.8\,\omega_{2jt-1} + \varsigma_{jt}\,, \tag{22a}$$

$$\varsigma_{jt} \sim i.i.d\ N(0,0.18)\,, \tag{22b}$$

In particular, we consider an AR(1) process with positive AR parameter, $\phi = 0.8$. This implies that the unmeasured product characteristics are now positively correlated over time and the shock is rather persistent. From (22a), we can also see that $Var(\omega_{2jt}) = Var(\varsigma_{jt})/(1-\phi^2) = 0.5$. Time-series plots of $\omega_{2jt}$'s (not shown for reasons of space) randomly generated from (22a)-(22b) confirm that $\omega_{2jt}$ is highly autocorrelated. The estimation model for our proposed method remains the same as in Case 1, leading to misspecification. Estimation results are shown in Table 1.3. We find that despite the misspecification estimates of all parameters are distributed around the true values, although dispersions of some parameters, , particularly $\overline{\beta}_1$ and $\overline{\beta}_2$, are larger than in Case 1.

**Table 1.3: Results of the Simulation Study Case 3 – Misspecification due to Autocorrelation in $\omega_{2jt}$ (based on 100 replications)**

| Parameters | True Values | T=50 | | T=100 | |
|---|---|---|---|---|---|
| | | Mean | SE | Mean | SE |
| $\overline{\beta}_1$ | 0.200 | 0.191 | 0.315 | 0.227 | 0.195 |
| $\overline{\beta}_2$ | 0.500 | 0.445 | 0.261 | 0.505 | 0.247 |
| $\overline{\beta}_3$ | -1.000 | -1.019 | 0.160 | -0.973 | 0.153 |
| $\overline{\beta}_4$ | 1.000 | 1.044 | 0.188 | 0.969 | 0.147 |
| $\theta_1 = \theta_2$ | 1.000 | 1.043 | 0.173 | 0.985 | 0.143 |
| $\sigma_{33}$ | 1.000 | 1.045 | 0.224 | 0.994 | 0.136 |
| $SD(\omega_{2jt})$ | 0.707 | 0.676 | 0.138 | 0.675 | 0.086 |

**1.4.4 Case 4: Misspecification of Distribution of $\omega_{1jt}$ and $\omega_{2jt}$**

As discussed, several pricing behaviors are inconsistent with the assumption of joint normality of the error in the pricing equation and the unmeasured product characteristics. In this study, we investigate the performance of the proposed method under non-Normal $\omega_{1jt}$. In particular, we generate $\omega_{1jt}$ and $\omega_{2jt}$ from Uniform distributions. All other details of the DGP remain the same as in Case 1 except (22) which is changed as follows:

$$\zeta_{1jt}, \; \zeta_{2jt} \sim i.i.d \; N(0, 0.5), \tag{22c}$$

$$\omega_{1jt}, \; \omega_{2jt} \sim Unif(-1,1), \tag{22d}$$

The estimation model for our proposed method remains the same as in Case 1, leading to misspecification.

Estimation results are shown in Table 1.4. Essentially, all the favorable results for the proposed model that were obtained in Cases 1-3 are retained under this form of misspecification.

**Table 1.4: Results of the Simulation Study Case 4 – Misspecification of the Distribution of $\omega_{1jt}$ and $\omega_{2jt}$ (based on 100 replications)**

| Parameters | True Values | T=50 | | T=100 | |
|---|---|---|---|---|---|
| | | Mean | SE | Mean | SE |
| $\bar{\beta}_1$ | 0.200 | 0.199 | 0.125 | 0.196 | 0.090 |
| $\bar{\beta}_2$ | 0.500 | 0.505 | 0.128 | 0.494 | 0.089 |
| $\bar{\beta}_3$ | -1.000 | -0.988 | 0.147 | -1.009 | 0.127 |
| $\bar{\beta}_4$ | 1.000 | 0.986 | 0.189 | 1.005 | 0.131 |
| $\theta_1 = \theta_2$ | 1.000 | 1.022 | 0.155 | 1.007 | 0.117 |
| $\sigma_{33}$ | 1.000 | 1.017 | 0.193 | 1.006 | 0.137 |
| $SD(\omega_{2jt})$ | - | 0.568 | 0.061 | 0.582 | 0.047 |

### 1.4.5 Case 5: Omitted Exogenous Unmeasured Product Characteristics

We examine the implications of ignoring unmeasured product characteristics even when they do not create an endogeneity issue. Chintagunta, Dubé, and Goh (2005) found that such omission led to higher estimated taste dispersion. However, we believe (and explain subsequently) that it is also possible that such misspecification would create biases in mean parameters. The specific data generating process (DGP) and the parameter values we assign are summarized below:

$$u_{hjt} = x'_{jt}\beta_h + \omega_{jt} + \varepsilon_{hjt}, \quad u_{h(J+1)t} = \varepsilon_{h(J+1)t}, \text{ if no purchase}, \quad \varepsilon_{hjt} \sim i.i.d. \; EVI, \quad (24)$$

$$x_{jt} = [c_{1jt} \quad c_{2jt} \quad x_{1jt} \quad x_{2jt}]',$$

$$c_{1jt} = I(j=1), \quad c_{2jt} = I(j=2),$$

$$x_{1jt}, \; \omega_{jt} \sim i.i.d \; N(0,1),$$

$$x_{2jt} = I(\upsilon_{jt} > 0.7), \quad \upsilon_{jt} \sim Unif(0,1),$$

$$\beta_h = \begin{bmatrix} \bar{\beta}_1 \\ \bar{\beta}_2 \\ \bar{\beta}_3 \\ \bar{\beta}_4 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_{33} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \eta_h \qquad \eta_h = \begin{bmatrix} 0.2 \\ 0.5 \\ -1 \\ 1 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \eta_h, \quad \eta_h \sim N(0, I_4),$$

where $I(\cdot)$ denotes indicator function and $I_k$ denotes a $k$-dimensional identity matrix. Note that $x'_{jt}\beta_h$, $\omega_{jt}$ and $\varepsilon_{hjt}$ are mutually uncorrelated. We assume that consumers are only heterogeneous in their taste for $x_{1jt}$, $J=2$, $T=100$ and 200, and $H=100,000$. We estimate the proposed model and a benchmark model that omits the exogenous unmeasured product characteristics, $\omega_{jt}$. The reason we consider larger values of $T$ in this simulation as compared with Cases 1-4 is to illustrate more precisely the biases this omission causes.

**Table 1.5: Results of the Simulation Study Case 5 – Exogenous Unmeasured Product Characteristics (based on 100 replications)**

| Parameters | True values | T=100 | | | | T=200 | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | | Proposed method | | Omitted $\omega_{jt}$ | | Proposed method | | Omitted $\omega_{jt}$ | |
| | | Mean | SE | Mean | SE | Mean | SE | Mean | SE |
| $\bar{\beta}_1$ | 0.200 | 0.212 | 0.192 | 0.310 | 0.115 | 0.230 | 0.129 | 0.301 | 0.086 |
| $\bar{\beta}_2$ | 0.500 | 0.495 | 0.170 | 0.551 | 0.099 | 0.471 | 0.128 | 0.540 | 0.073 |
| $\bar{\beta}_3$ | -1.000 | -0.951 | 0.144 | -0.790 | 0.113 | -0.946 | 0.119 | -0.817 | 0.100 |
| $\bar{\beta}_4$ | 1.000 | 0.961 | 0.221 | 0.807 | 0.148 | 0.949 | 0.195 | 0.803 | 0.119 |
| $\sigma_{33}$ | 1.000 | 0.984 | 0.190 | 0.808 | 0.213 | 1.004 | 0.169 | 0.839 | 0.167 |
| $SD(\omega_{jt})$ | 1.000 | 1.056 | 0.099 | - | - | 1.059 | 0.068 | - | - |

The first two moments of the empirical distributions for the parameter estimates are summarized in Table 1.5. The performance of the proposed method is similar to that in Cases 1-4. Estimates are tightly distributed around the true value. Considering the estimates when $\omega_{jt}$ is omitted from the estimation (see the relevant columns in Table 1.5), we see that when $T=100$, estimates of $\bar{\beta}_3$, $\bar{\beta}_4$, and $\sigma_{33}$ are biased toward zero. In particular, the true value of $\bar{\beta}_3$ is outside its 1.645 standard-error confidence band obtained from the empirical distribution of the parameter estimates (i.e. 90% confidence band of normal distribution). When $T$ becomes 200, the biases become more apparent. Now the true values for $\bar{\beta}_3$ and $\bar{\beta}_4$ are out of their 1.645 standard-error confidence bands obtained from the empirical distributions of the parameter estimates.

To explain these biases, let us rewrite equation (24) as $u_{hjt} = \mu_{hjt} + \omega_{jt} + \varepsilon_{hjt}$. If the exogenous unmeasured product characteristics $\omega_{jt}$ are ignored in the estimation, $\mu_{hjt}$ will absorb some part of the variation in $\omega_{jt}$ and $\varepsilon_{hjt}$ will absorb the rest. Consequently, we get $u_{hjt} = \tilde{\mu}_{hjt} + \tilde{\varepsilon}_{hjt}$ where $\tilde{\mu}_{hjt}$ and $\tilde{\varepsilon}_{hjt}$ are the new explained utility and unexplained utility, respectively, both inflated by the ignored $\omega_{jt}$. Since $Var(\tilde{\mu}_{hjt}) \geq Var(\mu_{hjt})$, we can expect over-estimated mean parameters and/or heterogeneity parameters. Similarly, we can expect $Var(\tilde{\varepsilon}_{hjt}) \geq Var(\varepsilon_{hjt})$ but this influences the model parameters in a way rather different from that of $\tilde{\mu}_{hjt}$. Due to the logit specification, the model will regard $\tilde{\varepsilon}_{hjt}$ as a random shock with a Type-I Extreme Value distribution and normalize the utility according to $Var(\tilde{\varepsilon}_{hjt})$. Since $Var(\tilde{\varepsilon}_{hjt}) \geq Var(\varepsilon_{hjt})$, $\tilde{\mu}_{hjt}$ and $u_{hjt}$ will be scaled down by the factor $Var(\varepsilon_{hjt})/Var(\tilde{\varepsilon}_{hjt})$, which is less than 1[8]. Consequently, we expect under-estimated mean parameters and/or heterogeneity parameters. Note that ignoring the EUPC

---

[8] This argument is similar to *attenuation bias* mentioned by Yatchew and Griliches (1985) in a homogeneous probit model.

influences the estimates of heterogeneity and mean parameters in two conflicting ways simultaneously: upward bias via $\tilde{\mu}_{hjt}$ and downward bias via $\tilde{\varepsilon}_{hjt}$. Since it is hard to predict which effect is greater, we cannot draw a general conclusion on the direction and severity of the biases.

In our simulated data it appears that the majority of the variance related to the ignored EUPC $\omega_{jt}$ is absorbed into the unexplained part of utility and consequently, the downward bias via $\tilde{\varepsilon}_{hjt}$ overwhelms the upward bias via $\tilde{\mu}_{hjt}$. Although not significant, we observe biases in the same direction in heterogeneity parameters and these can be interpreted similarly.

## 1.5 Empirical Application

### 1.5.1 Data

The data used in the study are histories of paper towel purchases of 880 households at an independent supermarket in Pittsburgh, Pennsylvania over 103 weeks through 1998 and 1999. The data are collected using a frequent shopper card. We include the four largest brands in the analysis: Bounty, Brawny, Scott, and Sparkle. The sales from these four major brands accounted for 77% of total category sales in our sample. Additionally, we include a "No purchase" option defined as shopping visits when none of the four brands of paper towels is purchased. Within each brand, the purchase of any one of different package sizes (i.e. number of rolls) was counted as a purchase of the brand. Price is defined on a per roll basis in our analysis. Price and promotion variables at the brand level were computed as market share-weighted averages of brand-size level variables. Descriptive statistics of the purchases, marketing mix variables, and wholesale prices, are provided in Table 1.6. About seven percent of store visits result in purchases of paper towels. Bounty is the dominant brand in the market, with over two-thirds market share.

**Table 1.6: Descriptive statistics of the Paper Towel Data**

**(Number of households = 880, Number of weeks = 103, Number of trips = 60,393)**

| Brands | Number of Purchases | Shelf Price ($ per roll) | | Wholesale Price ($ per roll) | | Correlation between shelf price and wholesale price | Promotion (% of store weeks) |
|--------|------|------|------|------|------|------|------|
| | | Mean | SD | Mean | SD | | |
| Bounty | 4,348 | 1.47 | 0.17 | 1.27 | 0.17 | 0.90 | 3% |
| Brawny | 781 | 1.56 | 0.34 | 1.37 | 0.31 | 0.92 | 11% |
| Scott | 558 | 1.39 | 0.39 | 1.13 | 0.28 | 0.94 | 6% |
| Sparkle | 778 | 1.00 | 0.13 | 0.79 | 0.10 | 0.54 | 29% |

To control for the endogeneity of price, we use weekly wholesale prices as instruments. (Following Kuksov and Villas-Boas (2003) and Chintagunta, Dubé, and Goh (2005), we assume that price is the only endogenous variable.) As expected, wholesale price is highly correlated with shelf price, with correlation coefficients ranging from 0.54 to 0.92 across the four brands. However, we do not expect the unmeasured product characteristics, especially those determined at retail (e.g. shelf space allocation) to be systematically related with wholesale prices. To the extent that this expectation is true, our instrumental variable is valid for controlling for the endogeneity of price. Brawny has the highest shelf and wholesale prices, followed by the largest brand Bounty. Brawny and Scott show high variances in both shelf prices and wholesale prices. Sparkle, the lowest-priced alternative, shows much more frequent promotion than other brands.

**1.5.2 Estimation and results**

Although our interest in this paper is in the aggregate model, we first estimate disaggregate models to obtain a benchmark (recall that the data are available at the household level). Disaggregate data contain complete information while aggregate data lose some information due to aggregation. Therefore, we expect that estimates from disaggregate data are more reliable than the ones from aggregate data for the same model specification. For the estimation of disaggregate models we use a method proposed by Chintagunta, Dubé, and Goh (2005) that we term ML/IV and describe subsequently.

Next, we aggregate our data to weekly brand sales data to estimate the proposed model as well as two other aggregate models –RCL and OEUPC. RCL is an aggregate version of the usual Random Coefficient Logit model. Therefore, it assumes that there is neither endogeneity of price nor unmeasured product

characteristics. Model OEUPC omits the exogenous unmeasured product characteristics, $\omega_{2,jt}$, as in simulation Case 5.

Using disaggregate data, we estimate the ML/IV model, which is a generalized two-step estimator suggested by Chintagunta, Dubé, and Goh (2005): 1) first, estimate the usual random coefficient logit model treating $\delta_{jt} = x'_{jt}\overline{\beta} + \xi_{jt}$ as fixed-effects and get estimates of fixed effects $\hat{\delta}_{jt}$ and two heterogeneity parameters, $\sigma_{price}$ and $\sigma_{promotion}$ [9]; 2) second, apply instrumental variable technique to $\hat{\delta}_{jt}$ considering the estimation error of the first step. Specifically,

$$\hat{\overline{\beta}}_{ML/IV} = (X'P_{\tilde{z}}\hat{\vartheta}^{-1}P_{\tilde{z}}X)^{-1}X'P_{\tilde{z}}\hat{\vartheta}^{-1}P_{\tilde{z}}\hat{\delta}$$ where $P_{\tilde{z}} = Z(Z'Z)^{-1}Z'$ is a projection matrix of instrument variable matrix $Z$, and $\hat{\vartheta}$ is covariance matrix of $\hat{\delta}$ from the first-step estimation. Using the disaggregate data we also estimate a generalized least-square estimator ML/LS: $\hat{\overline{\beta}}_{ML/LS} = (X'\hat{\vartheta}^{-1}X)^{-1}X'\hat{\vartheta}^{-1}\hat{\delta}$. ML/LS does not consider endogeneity of price and thus, we can investigate the bias due to endogeneity by comparing ML/IV and ML/LS.

Table 1.7 reports estimation results of three models with aggregate data (the proposed model, RCL, and OEUPC) and two models with disaggregate data (ML/IV and ML/LS).

---

[9] We assume that individuals have heterogeneous tastes with respect to price and promotion only. Our data provide nine moment conditions at most and, given these, we cannot entertain more complicated specifications of the heterogeneity distribution.

**Table 1.7: Estimation Results of Paper Towel Data**

| | Disaggregate Data | | | | | | Aggregate Data | | | |
| Model | ML/LS | | ML/IV | | RCL | | OEUPC | | Proposed method | |
| Parameters | Coef | SE | Coef | SE | Coef | SE | Coef | SE | Coef | SE |
|---|---|---|---|---|---|---|---|---|---|---|
| Bounty | 0.76 | 0.12 | -0.25 | 0.1 | 2.7 | 0.37 | 0.03 | 0.39 | -0.18 | 0.43 |
| Brawny | -0.65 | 0.13 | -2.25 | 0.11 | 0.98 | 0.36 | -1.88 | 0.39 | -2.16 | 0.44 |
| Scott | -1.35 | 0.12 | -2.62 | 0.1 | 0.2 | 0.34 | -2.19 | 0.37 | -2.47 | 0.41 |
| Sparkle | -1.7 | 0.1 | -3.85 | 0.08 | -0.53 | 0.31 | -2.97 | 0.34 | -3.28 | 0.39 |
| Price | -2.57 | 0.1 | -1.98 | 0.08 | -5.73 | 0.58 | -2.66 | 0.5 | -2.53 | 0.53 |
| Promotion | 1.22 | 0.19 | 1.55 | 0.14 | 1.12 | 0.16 | 1.25 | 0.1 | 1 | 0.52 |
| $\sigma_{Price}$ | 1.26 | 0.03 | 1.26 | 0.03 | 2.51 | 0.29 | 1.38 | 0.24 | 1.44 | 0.25 |
| $\sigma_{Promotion}$ | 2.11 | 0.14 | 2.11 | 0.14 | 1.49 | 0.22 | 0.31 | 0.32 | 1.5 | 0.64 |
| $b_{22,Bounty}$ | | | | | | | | | 0.27 | 0.03 |
| $b_{22,Brawny}$ | | | | | | | | | 0.74 | 0.07 |
| $b_{22,Scott}$ | | | | | | | | | 0.42 | 0.07 |
| $b_{22,Sparkle}$ | | | | | | | | | 0.72 | 0.09 |
| $\theta_{Bounty}$ | | | | | | | -1.87 | 0.27 | -1.64 | 0.41 |
| $\theta_{Brawny}$ | | | | | | | -4.66 | 0.39 | -4.45 | 0.49 |
| $\theta_{Scott}$ | | | | | | | 0.25 | 0.36 | 0.3 | 0.53 |
| $\theta_{Sparkle}$ | | | | | | | -6.39 | 0.34 | -6.57 | 1.12 |
| | | | | | | | | | | |
| AIC | | | | | 52,353 | | 52,241 | | 51,474 | |
| BIC | | | | | 52,386 | | 52,291 | | 51,542 | |
| log-like. | | | | | -26,168 | | -26,108 | | -25,721 | |

The disaggregate data results show a significant decrease in the absolute value of the price coefficient after using an instrumental variable. A Hausman test for price endogeneity rejects the null hypothesis of no endogeneity (i.e., $\bar{\beta}_{ML/LS} = \bar{\beta}_{ML/IV}$ ). The upward bias due to ignoring the endogeneity of price is contrary to typical previous findings of a downward bias (e.g., Besanko, Gupta, and Jain 1998; Chintagunta, Dubé, and Goh 2005, and others). In the literature, the unmeasured product characteristics are usually believed to be positively correlated with price, which is consistent with the commonly reported downward bias in the price coefficient when this correlation is ignored. However, our result implies a negative correlation between price and unmeasured product characteristics. One explanation for this negative correlation lies in expanded shelf space allocation or favorable shelf locations of price-promoted products, activities that are unobserved in our data. If shelf space changes are a dominant component of the unmeasured product characteristics, we can expect these characteristics to be negatively correlated with prices.

Turning next to results from aggregate models, all estimates from the RCL model are significantly different from those from ML/IV, which may be considered closest to the true parameter values. These "biases" are due to the omission of the unmeasured product characteristics as well as the ignored endogeneity. OEUPC estimates tell us what happens if we ignore the exogenous unmeasured product characteristics. Estimates of "Sparkle", "Promotion", and "$\sigma_{Promotion}$" are smaller in absolute value than the estimates from ML/IV. This pattern is similar to what we observed in Case 5. We may conclude that the majority of the variance related to the ignored exogenous unmeasured product characteristics $\omega_{2,jt}$ is absorbed into the unexplained part of utility and consequently, the downward bias via $\tilde{\varepsilon}_{hjt}$ overwhelms the upward bias via $\tilde{\mu}_{hjt}$ .

Estimates from the proposed method are all reasonably close to those of ML/IV and formally, they are not significantly different at a 5% significance level. We observe that standard errors of the proposed method are on average approximately 5 times those of ML/IV. This lower efficiency can be attributed to the information loss due to data aggregation. Estimate of "$\sigma_{Promotion}$" is significantly different from that of OEUPC. Moreover, estimates of $b_{22,Bounty}$, $b_{22,Brawny}$, $b_{22,Sparkle}$, and $b_{22,Scott}$ are all significant. These results imply that EUPC needs to be accommodated appropriately in the estimation.

By examining estimates of $\theta$'s, we can easily perform a formal test of endogeneity. Furthermore, the estimate indicates the sign of correlation between the unmeasured product characteristics and price.[10] We obtained highly significant negative estimates of $\theta_{Bounty}$, $\theta_{Brawny}$, and $\theta_{Sparkle}$, indicating that the endogeneity problem does exist in Bounty, Brawny, and Sparkles, and that the unmeasured product characteristics are negatively correlated with price. This confirms findings from the disaggregate models. The estimate of $\theta_{Scott}$ is not significantly different from zero and thus, we cannot reject the null hypothesis of no endogeneity in Scott. Using the previous display and shelf space allocation argument, we may attribute this to a different shelf allocation practice of the retailer with respect to Scott.

**1.6 Conclusion**

In this paper, we propose a Simulated Maximum Likelihood estimation method for the random coefficient logit model using aggregate data, accounting for heterogeneity and endogeneity. Our approach is suitable when observed brand shares contain sampling error. We show in simulated data that the proposed method provides unbiased and efficient estimates of demand parameters. Further methodological advantages of the proposed method include: 1) the proposed method provides

---

[10] Recall that $\theta'_j = b_{21,j} b_{11,j}^{-1}$ and $Cov(v_{jt}, \xi_{jt}) = b_{21,j}$.

endogeneity test statistics as a by-product; 2) it directly provides the direction of endogeneity bias, or the sign of correlation between endogenous regressors and the unmeasured product characteristics; and 3) the proposed method can be extended to incorporate Markov regime-switching dynamics in parameters and is open to other extensions based on ML.

Substantively, we also provide a more complete picture of the problems related to the omission of unmeasured product characteristics. We have shown that, in addition to the endogeneity problem, the omission can cause downward or upward biases in the estimates of mean parameters and/or heterogeneity parameters in the random coefficient logit model. Previously, Chintagunta, Dubé, and Goh (2005) noted an upward bias in heterogeneity parameters due to this omission and confirmed this result using scanner panel data on margarine purchases. In this paper, we identify the possibility of other biases. As an example, in simulation Case 5 we found that downward biases in mean parameters can result from an omission of the unmeasured product characteristics. This result was also found in an empirical application to paper towels data.

In the paper towel data, we also found a negative correlation between the unmeasured product characteristics and prices, a result that was confirmed by the disaggregate data. This finding is new to the literature. Furthermore, so far the correlation between price and the unmeasured product characteristics has been indirectly inferred from the direction of the endogeneity bias rather than directly estimated as in our proposed method.

The most important limitation of our proposed method is that the assumption of joint normality of the unmeasured product characteristics and the error in the pricing equation is inconsistent with a number of pricing behaviors. However, simulation experiments showed that misspecification of the distribution does not

hamper performance of the proposed method severely. We should also note that the proposed method is based on SML which is equivalent to ML when sufficient draws are used for the numerical integration. An attractive property of MLEs is that even under model misspecification, MLEs are strongly consistent in that they minimize the Kullback-Leibler Information Criterion. While this attractive property of MLEs alleviates our concerns somewhat, more rigorous study is needed with respect to misspecification of the unmeasured product characteristics.

The proposed method does not explicitly model serial correlation in the unmeasured product characteristics. However, simulation experiments showed that the autocorrelation in the unmeasured product characteristics does not hamper performance of the proposed method severely. We expect that a large part of this serial correlation can be captured by allowing Markov regime-switching dynamics in brand specific constants. Otherwise, we can extend the proposed method to explicitly model serial correlation by incorporating ARMA models. Extensions in this direction may also provide us with a deeper understanding of the unmeasured product characteristics.

REFERENCES

Albuquerque, P., B. J. Bronnenberg. 2006. Measuring consumer switching to a new brand across local markets. Working paper, University of Rochester, Rochester, NY.

Berry, S. 1994. Estimating Discrete-choice Models of Product Differentiation. RAND Journal of Economics. 25 242–262.

Berry, S., J. Levinsohn., A. Pakes. 2004. Differentiated Products Demand System from a Combination of Micro and Macro data: The New Car Market. Journal of Political Economy. 112(1) 68-105.

Berry, S., J. Levinsohn, A. Pakes. 1995. Automobile Prices in Market Equilibrium. Econometrica 63 841–890.

Berry, S., O.B. Linton, A. Pakes. 2004. Limit Theorems for Estimating the Parameters of Differentiated Product Demand Systems. Review of Economic Studies. 71(248) 613-54.

Besanko, D., S. Gupta., D. Jain. 1998. Logit Demand Estimation under Competitive Pricing Behavior: An Equilibrium Framework. Management Science. 44(November) 1533–1547.

Bodapati, A., S. Gupta. 2004. The Recoverability of Segmentation Structure from Store-Level Aggregate Data. Journal of Marketing Research. 41. 351–64.

Bronnenberg, B., V. Mahajan. 2001. Unobserved Retailer Behavior in Multimarket Data: Joint Spatial Dependence in Market Shares and Promotion Variables. Marketing Science. 20(3) 284-299.

Chen, Y., S.Yang 2007. Estimating Disaggregate Models Using Aggregate Data Through Augmentation of Individual Choice. Journal of Marketing Research. 44. 613-21.

Chintagunta, P. K. 2001. Endogeneity and Heterogeneity in a Probit Demand Model: Estimation using aggregate data. Marketing Science. 20(4) 442–456.

Chintagunta, P. K., J.-P. Dubé, K. Y. Goh. 2005. Beyond the Endogeneity Bias: The Effect of Unmeasured Brand Characteristics on Household-Level Brand Choice Models. Management Science, 51(5), 832–849

Draganska, M., D. C. Jain. 2002. Structural Models of Competitive Market Behavior. Management Science. 50 605–616.

Goolsbee, A., A. Petrin. 2004. The Consumer Gains from Direct Broadcast Satellites and the Competition with Cable TV. Econometrica. 72 351–381.

Hausman, J. 1978. Specification Tests in Econometrics. Econometrica. 46, 1251–1272.

Hausman, J. 1983. Specification and Estimation of Simultaneous Equation Models, in Handbook of Econometrics, ed. Z. Griliches and M. Intriligator, vol. 1, Amsterdam, North Holland.

Heckman, J. 1978. Dummy Endogenous Variables in a Simultaneous Equation System, Econometrica 46, 931–959.

Jiang, R., P. Manchanda, P.E.Rossi. 2007. Bayesian Analysis of Random Coefficient Logit Models Using Aggregate Data. Working paper, University of Chicago, Graduate School of Business.

Keane, M. P. 1993. Simulation Estimation for Panel Data Models with Limited Dependent Variables. C. R. Rao, G. S. Maddala, D. Vinod, eds. The Handbook of Statistics & Econometrics. Elsevier Science Publishers, Amsterdam, The Netherlands, 545–572.

Keane, M. P. 1997. Modeling Heterogeneity and State Dependence in Consumer Choice Behavior. Journal of Business and Economic Statistics. 15, 310–327.

Kitamura, Y., M. Stutzer. 1997. An Information-Theoretic Alternative to Generalized Method of Moments Estimation. Econometrica, Vol. 65, No. 4., pp. 861-874

Kuksov, D., J. M. Villas-Boas. 2007. Endogeneity and Individual Consumer Choice. forthcoming. Journal of Marketing Research.

Manchanda, P., P. Rossi, P. Chintagunta. 2004. Response Modeling with Nonrandom Marketing Mix Variables. Journal of Marketing Research, 41, 467-478

Musalem, A., E. T. Bradlow, J. S. Raju. 2007. Bayesian Estimation of Random-Coefficients Choice Models using Aggregate Data. forthcoming. Journal of Applied Econometrics.

Nevo, A. 2000. A Practitioner's Guide to Estimation of Random Coefficients Logit Models of Demand. Journal of Economics & Management Strategy, 9(4), 513-548.

Petrin, A. 2002. Quantifying the Benefits of New Products: the Case of the Minivan, Journal of Political Economy. 110, 705–729.

Petrin, A., K. Train. 2004. Omitted Product Attributes in Discrete Choice Models. Working paper, University of Chicago, Chicago, IL.

Shim, E., E. Sudit. 1995. How Manufacturers Price Products. Management Accounting. 76, 37-39.

Sudhir, K. 2001 a. Competitive Pricing Behavior in the US Auto Market: A Structural Analysis. Marketing Science. 20(1), 42-60.

Sudhir, K. 2001 b. Structural Analysis of Competitive Pricing in the Presence of a Strategic Retailer. Marketing Science. 20(3), 244-64.

Train, K., 2003. Discrete Choice Methods with Simulation. Cambridge University Press, Cambridge, UK.

Villas-Boas, M. and R. Winer. 1999. Endogeneity in Brand Choice Models. Management Science. 45, 10, 1324-1338.

Villas-Boas, M., Y. Zhao. 2005, Retailer, Manufacturers, and Individual Consumers: Modeling the Supply Side in the Ketchup Marketplace. Journal of Marketing

Research. Vol. 42 Issue 1, p83-95.

White, H. 1982. Maximum Likelihood Estimation of Misspecified Models.
Econometrica. 50, 1-25.

Wooldridge, J. M. 2001. Econometric Analysis of Cross Section and Panel Data.
Cambridge,
Mass.: MIT Press.

Wu, D. 1973. Alternative Tests of Independence Between Stochastic Regressors and
Disturbances. Econometrica. 41, 733-750.

Yang, S., Y. Chen, G. Allenby. 2003. Bayesian analysis of simultaneous demand and
supply. Quantitative Marketing and Economics. 1(3), 244–264.

Yatchew, A., and Z. Griliches. 1985. Specification error in probit models. The Review
of Economics and Statistics, 67, 134-139.

CHAPTER 2

PREFERENCE EVOLUTION IN THE SOUTH KOREAN CIGARETTE MARKET


## 2.1 Introduction

> *"Lower Tar Cigarettes the Choice of South Korean Smokers - With the*
> *'wellbeing' trend among consumers showing no signs of abating, smokers*
> *have flocked to lower tar products, and abandoned higher tar cigarettes in*
> *droves." - Euromonitor International*[11]
>
> *"We (KT&G) will be responsive to market demands, and continue to develop*
> *brands that meet consumers' increasing preference for low-tar and slim-type*
> *cigarettes." - KT&G Corporation*[12]

In the South Korean cigarette market, consumers have shown dramatic changes in
their preferences for cigarette attributes during the last decade.  For instance, while
most consumers smoked high-tar cigarettes ten years ago, now most consumers prefer
low-tar cigarettes.  According to KT&G, the leading cigarette manufacturer in South
Korea, the market share of cigarettes which contain less than 3 mg of tar per cigarette
was a mere 1.8% in 2002.  This number had grown to about 50% by 2007.  Another
interesting trend in this market is the growing popularity of super-slim cigarettes,
which are very thin cigarettes with a diameter of 5.4 mm in contrast to the 7.8 mm
diameter of regular cigarettes.  While in other countries super-slim cigarettes are
usually popular only with female smokers, in South Korea they have paradoxically
become the cigarette of choice of many male smokers and are now the best-selling
types of cigarettes in the South Korean market.

---

[11] "Tobacco in South Korea," Euromonitor International (www.euromonitor.com /Tobacco _in_ South_Korea)
[12] "2003 Annual Report," KT&G Corporation

Given these strong dynamics in preferences, I raise two critical questions – 1) What drives consumer preference changes, and 2) How does the firm react to these preference changes. I answer these questions by proposing a unique structural model of consumer demand and firm behavior. With respect to the first question, I classify possible sources of preference changes into two categories: exogenous effects and marketing effects. First, consumer preference changes can be driven by exogenous effects in the sense that they are not caused by the firm (e.g. social trends). In particular, the growing preference for low-tar cigarettes can be explained by consumers' increased health consciousness and their belief that smoking low-tar cigarettes is less harmful. A recent study reveals that the majority of smokers who switch to low-tar cigarettes believe that it will help them kick the habit or reduce their risk of tobacco-related diseases (Shin 2007).[13] Along with the exogenous effect, marketing activities may also influence consumer preferences (e.g. Mela et al. 1997). However, in the cigarette market the "usual" marketing instruments are restricted or controlled by the government. In South Korea, advertising and promotion of cigarette products is largely prohibited and pricing is under strong government restriction. However, the government seems to be less concerned with the introduction of new products and introductory prices. Since the new product may convey critical information to consumers, I believe this serves as an important marketing instrument. Consequently, I focus on the influence of new product introduction on consumer preference evolution. I propose an aggregate random coefficient logit model where the parameters evolve as a function of the introduction effect and the exogenous effect.

---

[13] In the U.S., a similar smoker transition from high-tar to low-tar products has been reported. The switchers cited switching to low-tar cigarettes as a strategy to help them stop smoking and also reported that they believed low-tar cigarettes were safer than high-tar cigarettes. Source: "Low-Tar Cigs May Not Help Smokers Quit," WebMD (www.webmd.com/smoking-cessation/news/ 20031023/ low-tar-cigarettes-smokers-quit).

This model allows me to separate one effect from another and examine the significance of each effect.

Another key research question I study is how the firm reacts to consumer preference changes. To answer this question, I build two supply side models – a pricing model and a new product introduction and attribute selection model. First, I specify the firm's pricing model, which elucidates the influence of the time-varying preferences on the firm's price decisions. As in many other countries, the government regulates cigarette prices as well as other marketing mix variables in this market. Thus, the market prices are not simply set at the profit maximizing levels. Considering this restriction, I propose a flexible pricing model which does not impose any equilibrium condition. Second, I model the firm's decisions regarding new product introduction and selection of attribute-levels for product design. In each time period, the firm can choose to not introduce any product, or to introduce a product from a set of candidate products that are created as combinations of underlying product attributes.

In the data I observe 17 new product introductions during the 106-month sample period from January 1995 to October 2003. I model changes in consumer preference for product attributes as outcomes of both an exogenous effect and the effect of new product introductions during this period. As consumer preferences evolve over time, the optimal new product design and its profitability vary as well. This provides informative variations that help identify the firm's decision process regarding whether to introduce a new product, and its optimal design. Herein lies the uniqueness of my analysis compared to extant studies that analyze firms' new product introduction or product positioning decisions. To sum up, the market evolves through the interaction of consumer preference changes and firm actions and I examine this industry-wide dynamic in detail using the proposed model.

From a methodological perspective, I identify new endogeneity issues that arise due to time varying preferences. Empirical results show that consumer preferences are correlated with both, cigarette prices and the attribute levels of products in the market over time. If ignored, these correlations cause serious biases in the estimated parameters. Another methodological contribution of this study is that the proposed demand model extends an aggregate random coefficient logit model to incorporate stochastically evolving parameters in a likelihood-based framework. Thus, the demand model can be jointly estimated together with the supply side models using MLE, leading to efficiency. Moreover, structural tests can be easily performed using the standard likelihood-based hypothesis test procedures (e.g. Wald test, likelihood ratio test).

The main implications of the estimation result are the following: (1) consumer preferences for cigarette attributes have significantly changed during the sample period; (2) the firm strategically sets cigarette prices, taking advantage of the time-varying nature of consumer preferences; (3) preference changes play an essential role in explaining observed variations in product attributes of newly introduced products, in other words, the firm's new product design and introduction strategy.

The main contributions of this study will be the following:

i.   It provides a complete framework of market evolution which is modeled as an outcome of the interaction between the firm's marketing activities and changes in consumer preferences. My empirical results will provide insights into how the preference changes and the firm decisions shaped the fundamentals of the cigarette market in South Korea.

ii.  This study sheds light on the role and value of new product design and introduction in a market where consumer preferences are evolving, and the firm's ability to influence demand through prices or advertising is constrained

by regulation. The proposed model will be used to perform counterfactual analyses to assess how the market would have evolved if the firm had not introduced the new products that we observe in the data.

iii. The proposed model can help the firm determine a new product strategy that will direct the market in a direction that is preferred by the firm. Using policy simulations, the firm can compare profits under alternative new product introduction policies.

iv. Methodologically, this study identifies new endogeneity problems due to time-varying consumer preferences. Also, it extends the aggregate random coefficient logit model to incorporate stochastically varying parameters in a likelihood-based framework.

v. The findings of this study will provide rich implications for governments and allow them to build effective policy. The price elasticity of cigarette demand has been extensively studied in the cigarette-related literature and has played a prominent role in legislative debates about using taxation as a principal tool to discourage smoking. If consumer preferences are changing over time, proper consideration of this is necessary for the precise estimation of the price elasticity. I show the importance of this aspect using real data.

The remainder of this paper proceeds as follows. In section 2.2.2, I provide details on the South Korean cigarette market and an overview of the relevant literature. In section 2.2.3, I describe the data sets used in this study and results of preliminary analyses. The goal of the preliminary analysis is to understand the important aspects of the data thereby enriching the proposed model. In section 2.2.4, I present the model which will be applied to data from the South Korean cigarette market. The model consists of three sub-models; 1) demand model, 2) pricing model, and 3) new product introduction and attribute choice model. I also provide the details of estimation

procedure. In section 2.2.5, I present the results and discuss the implications. The conclusion and future direction of research follow.

## 2.2 Background and Related Literature

### 2.2.1 Tobacco Industry

In many countries, the tobacco industry ranks among the most substantial and successful of economic enterprises (Chaloupka and Warner 2000). The most influential player in this industry is oftentimes the government. Government control of tobacco through the creation of a monopoly, for instance, has existed in Austria, Spain, Portugal, France, Italy, and Germany. In Asia, the domestic tobacco industries of South Korea, China, Japan and several other countries are, or have been at one time, state-owned enterprises.[14] To governments, tobacco is an important matter, touching on a variety of critical issues including politics, agriculture, employment, trade, and tax revenue. Among these, tax revenue may have primary importance. To this day tobacco-related revenues are a welcome source of cash for many governments. In South Korea and Japan, approximately 2% of total tax revenue comes from tobacco related taxes. Tobacco taxes represent about 3% of the government's total revenue in the U.K., France, and Germany. In China, approximately 10% of total tax revenue comes from tobacco. While taxes from tobacco products are now a small percentage of total tax revenues in the U.S., in 1960 they represented 2.5% of total federal and state revenue (Feldman 2001).

The frequently observed highly concentrated nature (i.e. monopoly or oligopoly) of the tobacco industry can be explained by industry characteristics. Cigarette manufacturing belongs to a capital-intensive process industry, where economies of scale are key success factors and the gestation period of capital is

---

[14] Other countries that now have (or once had) tobacco monopolies include Ethiopia, Iceland, Jordan, Syria, and Zambia, among others.

relatively long.  It also requires large-scale purchase of raw materials (leaf tobacco) and an aging process that lasts for two years.  Considering all these factors, monopoly or oligopoly might very well be the optimal structure of the industry.

## 2.2.2 South Korean Cigarette Market

In South Korea, 99.7% of tobacco is consumed as cigarettes, all of which are filter-tipped.[15] Other tobacco variants remain unpopular, and have insignificant sales and distribution in the market.  Cigars are only available in specialty stores, usually of the Cuban variety and costly compared to cigarettes.  The availability of smokeless tobacco and pipe tobacco is extremely limited.[16] Therefore, the term "tobacco" is almost equivalent to "cigarette" in South Korea.

Following its establishment in 1948, the fledgling South Korean government took control of the cultivation, manufacture, and sales of tobacco products.  The government had control of the cultivation of leaf tobacco through the Leaf Tobacco Monopoly Law.  Also, cigarette factories throughout the country were owned and operated by the government.  Retailers, licensed by the government, sold their products at a fixed price, which was also pre-determined by the government.  In 1987, the government established the Korea Monopoly Corporation, a government-invested company, as an agency of the government's tobacco- and ginseng-related businesses.  In 1988, when the U.S.  pressured South Korea to open its market to foreign tobacco products, the government liberalized the import of non-Korean cigarettes and renamed the Korea Monopoly Corporation the Korea Tobacco and Ginseng Corporation.  Since then, the government's general policy with respect to the domestic tobacco industry has been consistently toward liberalization and privatization.  In 1997, the Korea Tobacco and Ginseng Corporation was converted from a state-owned enterprise into a

---

[15] "Economics of Tobacco Control," World Bank. (www.worldbank.org/tobacco/brieflist_db_print.asp)
[16] "Tobacco in south Korea," Euromonitor International. (www.euromonitor.com/ Tobacco _in_ South _Korea)

private company.  In the years that followed, it held a successful initial public offering

and was listed on the Korea Stock Exchange.  However, the government still remained

the dominant stock holder.  In 2002, the government sold its shares of the Korea

Tobacco and Ginseng Corporation, and the company was fully de-nationalized and

renamed the KT&G Corporation.[17]

Although the government's controlling role in tobacco has been diminished, it

still remains the most influential player in the industry and has been increasingly

interested in controlling the practice of smoking.  Anyone desiring to manufacture

and/or sell tobacco must still obtain a license from the government.  Also, the

government still has the power to determine cigarette prices and the tobacco-related

taxes,[18] restrictions on the advertising of cigarettes and smoking in public places, and

almost every other important aspect of the business.  Unlike many other countries,

cigarette advertising has not been a subject of policy debate since there has been little

advertising.  Beginning from the monopoly period, cigarette advertising has been

limited to short-lived print media campaigns to introduce new products and has been

banned on television and radio.[19]

By law, all cigarette retailers must charge the pre-determined prices and no

price-related promotion by a retailer is allowed.  Thus, there is no cross-sectional

variation in the retail price of a given cigarette product in South Korea.  Another

interesting feature of this market is that cigarette prices increase only when the

government increases the tobacco-related taxes.  As a result, cigarette prices over time

have the shape of a step function and can be divided into price regimes.  Figure 2.1

shows the price changes of selected brands over time.  There have been four price

---

[17] In this study, I use the term "KT&G" to include its predecessors (i.e. the Korea Tobacco and Ginseng Corporation and the Korea Monopoly Corporation).

[18] Tobacco-related taxes include tobacco excise tax, the value added tax, and dues.

[19] Print media are limited to magazines (excluding women's and children's magazine). Also, there are strict restrictions on frequency of advertising, spaces or pages, and contents.

increases (due to tax increases) between 1995 and 2003, as the step changes in the figure show.

Although non-Korean cigarettes (i.e. foreign cigarette manufacturers' products) have been sold since 1988, meaningful competition between Korean cigarettes (i.e. KT&G products) and non-Korean cigarettes is recent. KT&G had maintained over 90% of the domestic market share until 2000 (Min 2007). Thus, I regard KT&G as a monopolist during the sample period. Since the recent deregulation of the market, foreign cigarette manufacturers have been moving aggressively to claim a stake in the dominant market share hitherto held by KT&G. Along with KT&G, the South Korean cigarette market currently includes products of British American Tobacco (BAT), Philip Morris (PM) and Japan Tobacco International (JTI), and all three of these foreign cigarette manufacturers now have manufacturing facilities in South Korea. The recent intensification of competition has caused KT&G's market share to decline steadily, although it still holds the lion's share of 70% of the market.

## 2.2.3 Literature Review

New product introduction and product design are central marketing decisions. I classify the related literature into two streams: 1) decision support analysis and 2) structural analysis. The main goal of the first stream of research is to help marketers make better decisions regarding new product introduction and product design, or more broadly, product line decisions (e.g. Green and Krieger 1985; Dobson and Kalish 1988). On the other hand, the main goal of the second stream of research is to make inferences about firm or market primitives, assuming that agents are behaving in a certain manner, and that researchers observe the outcomes of agent interactions (e.g. Mazzeo 2002; Draganska et al. 2009).
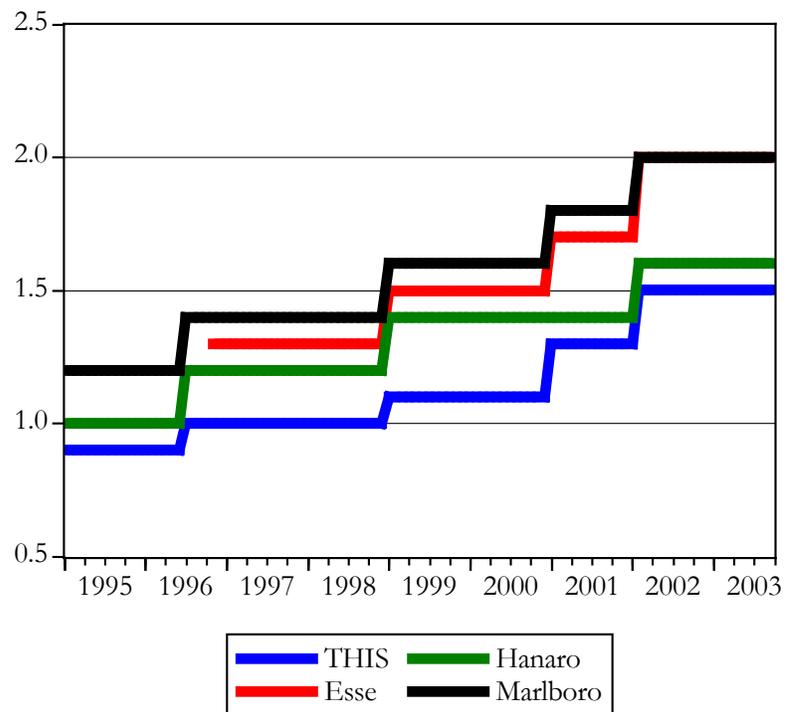
**Figure 2.1: Cigarette Prices of Selected Brands (x-axis: time; y-axis: 1,000 Korean Won which is approximately 1 US. Dollar)**

The main question addressed by the decision support analysis literature is: How should the firm position and price a line of related products in order to maximize profits or welfare? In the single product problem the optimal product design and price can often be obtained analytically (for a review, see Shocker and Srinivasan 1979). Later research considers the multi-product problem. For example, Green and Krieger (1985) consider the problem of determining the best subset of a given set of products to introduce, based on utility estimates of a sample of consumers. Dobson and Kalish (1988) extend this research by explicitly considering price as a separate attribute and incorporate a realistic cost structure into the modeling framework. Horsky and Nelson (1992) further extend the research by addressing the new product pricing and positioning problem in an oligopolistic market with competitor reactions.

Recent literature using structural models has made further progress on the new product introduction question. Kadiyali et al (1999) examine the impact of a line extension on price competition between two national yogurt manufacturers. Their result tells us that the extending firm gains price-setting power. Hitsch (2006) investigates how a firm in the US breakfast cereal market should make product launch and exit decisions when there is uncertainty about demand. The result implies that under some level of uncertainty, the firm should launch a new product even when it is expected to be unprofitable. This explains the high failure rate of the industry. Shen (2008) investigates the firms' entry and exit decisions during the formative period of a new industry. The proposed dynamic model provides an explanation for observed industry evolution patterns in terms of change of the number of firms, prices and total outputs. Draganska et al (2009) examine the product assortment decisions of oligopolistic firms by treating product choice as endogenous. The joint modeling of product assortment and pricing decisions improves standard product choice models by

allowing insights into how demand characteristics affect firms' product offerings in a competitive environment.

My research belongs to the second stream, namely, structural analysis. The distinguishing feature of my study is that consumer preferences for product attributes change over time. I develop a demand model that explicitly allows consumer preferences to vary over time, influenced by the firm's marketing activities (i.e. new product design and introduction) and other exogenous factors (e.g consumers' increasing health consciousness). To my knowledge, most extant models in the related literature are based on the premise that consumer preference is unchanging or fixed. I also examine the influence of the preference change on the firm's decision. I find that dynamics in preferences play a significant role in explaining the firm's choices of which products to introduce over time. I also investigate whether the firm varies prices to take advantage of changing preferences to enhance profits. Econometrically, the dynamics in consumer preferences raise a potential source of endogeneity biases in demand estimation. This concern arises if the firm designs (i.e., chooses attribute levels) and introduces new products to take account of changing consumer preferences. To summarize, my main research questions are all related with changing consumer preferences, and this makes my work relevant to a number of markets where evolution of consumer tastes is an important phenomenon. This aspect is also what distinguishes my research from extant literature.

From a methodological perspective, the proposed demand model extends the aggregate random coefficient logit (RCL) model to incorporate time-varying coefficients in a likelihood framework. The RCL model that is based on aggregate data and incorporates price endogeneity via instrumental variables is now a popular tool for the empirical analysis of demand in differentiated product markets (Ackerberg, Benkard, Berry, and Pakes 2007). Berry, Levinsohn, and Pakes (1995, henceforth

BLP) pioneered this model, proposing a Generalized Method of Moments (GMM) estimator for the estimation. Since then a number of alternative models and estimation methods have been proposed (e.g. Petrin and Train 2009; Park and Gupta 2009; Musalem et al. 2009; Jiang et al. 2009). Among these variants, a noteworthy approach is a likelihood-based approach. By imposing a distributional assumption on the common demand shocks (or unmeasured product characteristics), one can derive a likelihood function, which entails a Jacobian matrix corresponding to the transformation of variables of the common demand shocks to market shares. The estimation can be performed either in the classical framework or in the Bayesian framework. Jiang et al (2009) develop a Bayesian estimator. They report the advantages of the likelihood-based approach over the original GMM approach. Most of all, researchers can expect efficiency gain by applying the likelihood-based approach and the efficiency gain is preserved even under misspecification of the distribution of the common demand shock. Park and Gupta (2008) compare an alternative simulated maximum likelihood (SML) estimator to the standard GMM estimator of BLP. They also confirm the efficiency gain of the likelihood-based approach. In this paper, the proposed demand model extends the likelihood-based aggregate RCL model to incorporate time-varying coefficients. The Kalman filtering algorithm is applied to handle time-varying preferences and this fits nicely into the likelihood framework. A similar aggregate discrete choice model with time-varying parameters is proposed in a GMM framework by Sriram et al. (2006). I expect analogous advantages in the proposed likelihood-based model when compared to a GMM-based method.

One can easily modify the proposed model to accommodate a Markov switching (or Hidden Markov) process for the coefficients. For that, one can apply the Hamilton filter (i.e. the procedure used to estimate Markov switching part of the

52

model; Hamilton 1989) instead of the Kalman filter. Since I can write the complete likelihood-function of the model, I can readily develop a Bayesian estimator. Moreover, structural tests can easily be performed using the standard likelihood-based hypothesis test procedure (e.g. Wald test, likelihood ratio test). One promising extension of the model is to add a hierarchical structure to the aggregate RCL model. Let us assume that we have access to the market shares of multiple markets over time. We can think of a two-layered model: the lower-layer consists of the market-level aggregate RCL models and the upper-layer specifies the relationship between the market-level parameters. This model is analogous to the hierarchical Bayes discrete choice model. The proposed model can easily be modified to accommodate this.

In the marketing literature, there have been several studies regarding the U.S. cigarette market. Holak and Reddy (1986) explore the effects of the cigarette industry's television and radio advertising ban of 1970 on the price elasticity, advertising elasticity, and brand purchase inertia. After the ban, product demand becomes more price-sensitive and more inelastic with respect to advertising fluctuations. Also, brand purchase inertia becomes significantly higher. Chen et al (2007) empirically study the impact of Marlboro's permanent price cut in 1993 to stop the erosion in its market share resulting from the introduction of cheap generic brands. In particular, they examine whether a permanent price cut causes consumer choice behavior to shift over time using a dynamic structural brand choice model with learning and time-varying coefficients. The result indicates that the permanent price cut was effective in encouraging consumers to adjust their preferences to the new pricing policy and the newly established consumer preferences help alleviate erosion of Marlboro's market share. In economics, there exists a huge literature regarding smoking (for a review, see Chaloupka and Warner 2000). Cigarette price elasticity has been the most extensively studied topic in this literature and played a prominent

role in legislative debates about using taxation as a principal tool to discourage smoking. Another popular topic in this literature is the impact of advertising on cigarette consumption and demand. A related literature is on the theoretical and empirical modeling of the demand for addictive products (e.g. the rational addiction model of Becker and Murphy 1988; Machado and Sinha 2007).

Extant studies in both marketing and economics mainly focus on the consumer's response to changes in marketing mix variables (e.g. decreased advertising or price cut). In contrast, my study focuses on the interaction between the firm's actions (i.e. new product design, pricing and introduction) and the consumer's behavioral changes (i.e. changes in preference for cigarette attributes). In many countries, governments are increasingly concerned about smoking and public health. In the U.S., a bill that increases the power of the Food and Drug Administration (FDA) to regulate the manufacturing, marketing and sales of tobacco products recently cleared the Senate. In many countries, the cigarette industry's price, promotion and advertising activities have been regulated by the government. By examining firm behavior regarding new product design and introduction, this study will provide implications for new tools of regulation.

## 2.3 Data and Preliminary Analyses

In this section, I describe the data and results of a series of preliminary analyses. The goal of these analyses is to understand the important aspects of the data and, based on such knowledge, build a complete modeling framework to be presented in Section 2.2.4.

### 2.3.1 Data

I obtained data from KT&G Corporation on monthly sales, prices, and characteristics of 31 cigarette products in the South Korean market from January 1995 to October 2003 (106 months). All 31 products are manufactured by KT&G and

collectively represent 85% of total cigarette sales in the market.  Given its dominant

market share, I regard KT&G as a monopolist during the sample period and build the

supply side model accordingly.  At the beginning of the time window there are 14

products in the market, and at the end 27 products.  During the sample period, KT&G

introduced 17 new products and withdrew 4 products from the market.  Further, during

this period, there were four tax increases and related cigarette price increases (July

1996, January 1999, January 2001, and February 2002).  In January 1995, the share-

weighted average price of a pack of cigarettes (20 cigarettes) was 810 Korean Won

(hereafter, KRW; 1,000 KRW is approximately equivalent to 1 US Dollar); this

increased to 1,767 KRW in October 2003.  Thus, nominal prices more than doubled

over this time.

Figure 2.2 shows the changes in total cigarette sales in the market, aggregate

sales of products included in this study, and share-weighted average prices of included

products over time.  While the average price more than doubled, total sales do not

show any significant trend over time.  A simple regression of aggregate cigarette sales

on the average price (and a constant) reveals that the effect of price is insignificant.[20] I

obtain valuable additional information by viewing cigarettes as differentiated products,

a perspective that is popular both in marketing and in the empirical industrial

organization.  Therefore, I allow substitutions between products due to price changes

and thereby identify the price effect.  The results from all the models I consider show a

statistically significant negative effect of price on demand.  While there is a huge body

of literature in health economics and public policy on the effect of cigarette prices on

demand, the effect on product substitutions has largely been overlooked.  As my

findings suggest, a differentiated products perspective should add to our understanding

of the effects of cigarette price changes.

---

[20] The p-value of price coefficient is 0.2.
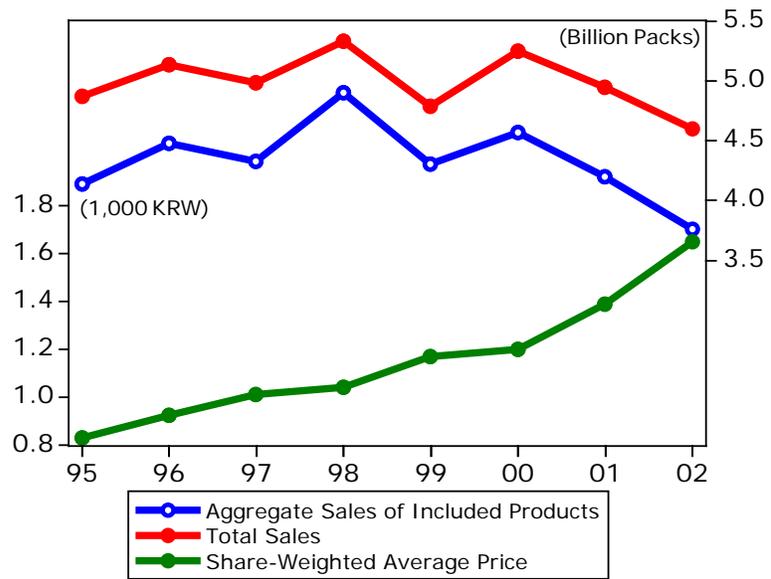
**Figure 2.2: Cigarette Sales and Share-Weighted Average Price**

## 2.3.2 Preliminary Analysis

I explore the data using homogenous logit models. While the homogenous logit model assumes restrictive and unrealistic substitution patterns, it is a useful tool to get a feel for the data. At each time period $t = 1,...,T$, the utility of alternatives $j=0,..., J_t$ for consumer $h=1,...,H$ is given by the following expression:

$$U_{h0t} = \widetilde{\xi}_{0t} + \varepsilon_{h0t}, \tag{1}$$

$$U_{hjt} = X'_j \beta + P_{jt}\alpha + Y'_{jt}\rho + \widetilde{\xi}_{jt} + \varepsilon_{hjt}. \tag{2}$$

$j=0$ denotes "no purchase" or "outside good." $P_{jt}$ is price of brand $j$ at $t$ (I use 1,000 KRW as the units here). A (6×1) vector of dummy variables $X_j = [X_j^{sl}\ X_j^{ss}\ X_j^{mt}\ X_j^{lg}\ X_j^{lt}\ X_j^{ht}]'$ contains information regarding the product attributes. In this study, I consider four product attributes: thickness, menthol taste, length, and tar content. The attribute thickness has three levels: regular ($X_j^{sl}=0$, $X_j^{ss}=0$), slim ($X_j^{sl}=1$, $X_j^{ss}=0$), and super-slim ($X_j^{sl}=0$, $X_j^{ss}=1$). Menthol taste has two levels: yes ($X_j^{mt}=1$) and no ($X_j^{mt}=0$). Length has two levels: regular (84 mm; $X_j^{lg}=0$) and long (100 mm or longer; $X_j^{lg}=1$). Finally, tar content has three levels: low (tar content is less than 5 mg; $X_j^{lt}=1$, $X_j^{ht}=0$), medium (tar content is 5 mg to 8 mg; $X_j^{lt}=0$, $X_j^{ht}=0$), and high (tar content is greater than 8 mg.; $X_j^{lt}=0$, $X_j^{ht}=1$).[21]

$Y_{jt}$ contains variables to capture the stock-piling behavior before the price increases (dummies for the two months before a price increase; denoted by *PIB1* and *PIB2*), temporary sales decrease after the price increases (dummies for the month of price increase and the next month; denoted by *PIA1* and *PIA2*), dummies for January and February, and dummies to capture the impact of special round pricing ($P_{1000} = I(P_{jt} = 1,000KRW)$, $P_{1500} = I(P_{jt} = 1,500KRW)$, $P_{2000} = I(P_{jt} = 2,000KRW)$, and $I(\cdot)$ is indicator function). The rationale for including these variables is as follows.

---

[21] I choose these cut-off values because they classify the sample products into three subgroups of the similar size. I also tried different cut-off values but the main findings from the result was not sensitive to the changes.

Price increases are usually announced in advance, thereby leading to possible

stockpiling effects. Many smokers decide to quit smoking at the start of the New Year.

However, the majority of them fail to quit and resume smoking in one or two months

and sales come back to the usual level. I capture this phenomenon using January and

February dummies, which I expect will have negative effects.[22] Some consumers seek

transactional convenience and prefer round-priced products since they save the effort

of keeping small coins to pay.[23] I also include terms to capture long term trends in the

life cycles of cigarette products. I model trend using a flexible function of $Age_{jt}$,

which is months elapsed since the introduction of the product. In particular, I use

$Age_{jt}$, $Age_{jt}^2$, and $1/Age_{jt}$ since these allow high flexibility in functional form while

maintaining parsimony. $\widetilde{\xi}_{jt}$ are unmeasured product characteristics (UPC) which may

include, for example, the impact of unobserved promotional activity, advertising,

unquantifiable factors and systematic shocks to demand. $\varepsilon_{b0t}$ and $\{\varepsilon_{bjt}\}_{j=1,...,J_t}$ are i.i.d.

random shocks with a type I extreme value distribution.

   Letting $\widetilde{\xi}_{0t} = 0$ for the normalization, I can derive choice probabilities from

(1)-(2). As in BLP, I assume that there is no sampling error in the aggregate data.

Then, the following linear equation can be derived:

$$\ln(s_{jt}/s_{0t}) = X'_j \beta + P_{jt} \alpha + Y'_{jt} \rho + \overline{\xi} + \xi_{jt}, \quad j = 1,...,J_t \tag{3}$$

where $s_{jt}$ and $s_{0t}$ denote market shares of product $j$ and of "no purchase" at $t$,

$\widetilde{\xi}_{jt} = \overline{\xi} + \xi_{jt}$ and $\overline{\xi}$ is the mean of $\{\widetilde{\xi}_{jt}\}_{j=1,...,J_t, t=1,...,T}$. $s_{jt}$ and $s_{0t}$ are calculated from

the sales of each product and market potential $M_t$.[24] I assume that $\xi_{jt}$ has a mean

zero and a finite variance $\sigma_{\xi}^2$ and from this I can obtain least-squares estimates of

model parameters. One major concern with the specification of (1)-(2) is the price

---

[22] This smoker behavior is also confirmed through the interviews with retailers and KT&G personnel. I also tried dummies for the other months but found they are not significant.
[23] This is also confirmed through interviews with retailers and KT&G personnel.
[24] I define market potential assuming that all individuals in the population over 18 (which is the legal smoking age in South Korea) smoke one pack (20 cigarettes) a day on each of 30 days in a month.

endogeneity. The UPC might be correlated with price. Many researchers have discussed mechanisms which account for the correlation between UPC and price (BLP 1995; Villas-Boas and Winer 1999).

In this study, I identify another source of possible endogeneity. If consumer preferences are changing over time but the empirical model does not appropriately incorporate this, then the impact of time-varying preference might be absorbed into $\xi_{jt}$. To make this precise, say that the true time-varying preference vector is $\beta_t$ but is erroneously modeled as $\overline{\beta}$. Then, the new UPC becomes $\breve{\xi}_{jt} = \xi_{jt} + X'_j(\beta_t - \overline{\beta})$. On the other hand, the firms know that consumers' preferences are changing and strategically set prices based on their knowledge of time-varying preferences. That is, $P_{jt} = f(\beta_t)$. Now it is clear that $\breve{\xi}_{jt}$ is correlated with $P_{jt}$ through $\beta_t$ and this, if ignored, results in the biased estimates of model parameters. To overcome potential problems due to endogeneity, I use instrumental variables (IVs). In particular, I use prices in other price-regimes as instruments. This is based on the assumption that prices in other price-regimes are uncorrelated with $\breve{\xi}_{jt}$.[25] To the extent that this belief is true, my IVs are valid to control for the correlation between the UPC and price. Using the IVs, I estimate the following equation:

$$\ln(s_{jt} / s_{0t}) = X'_j\beta + P_{jt}\alpha + Y'_{jt}\rho + \overline{\xi} + CF_{jt}\vartheta + \widehat{\xi}_{jt} \tag{4}$$

where $\breve{\xi}_{jt} = CF_{jt}\vartheta + \widehat{\xi}_{jt}$. $CF_{jt}$ is the residual from the regression of $P_{jt}$ on the IVs and works as a bias correction term[26], referred to in the literature as "control function" (Petrin and Train 2009). By introducing $CF_{jt}$, the new error term $\widehat{\xi}_{jt}$ becomes uncorrelated with any term on the right-hand side of (4). Consequently, the least square estimates of (4) are now unbiased.

---

[25] This assumption can be violated if $\beta_t$ is correlated over time. However, I use these IVs since it is difficult to obtain other suitable IVs. To alleviate the concern of autocorrelation, I also used prices in price regimes that are far from a particular $t$ and obtained a similar result.
[26] In linear models like the current case, the classical IV method is equivalent to eq. (4) which incorporates residuals from IV equation.

Along with the endogeneity issue, I examine whether consumer preferences change over time in the simple setup of (4). I split the data set into two halves and estimate separate coefficients for cigarette attributes and price for each sub-period. Let $\tau$ denote the midpoint in time in the data set. The estimating equation is,

$$\ln(s_{jt} / s_{0t}) = X'_j \beta + P_{jt} \alpha + Y_{jt} \rho + \bar{\xi} + CF_{jt} \vartheta +$$
$$X'_j \cdot \Delta\beta \cdot I(t > \tau) + P_{jt} \cdot \Delta\alpha \cdot I(t > \tau) + \Delta\bar{\xi} \cdot I(t > \tau) + \hat{\xi}_{jt} \qquad , \qquad (5)$$

where $\Delta\beta$ is a (6×1) coefficient vector, and $\Delta\alpha$ and $\Delta\bar{\xi}$ are scalar coefficients.

Table 2.1 shows the estimation results of three models: (3), (4) and (5). By comparing (3) and (4), I can examine whether the endogeneity problem exists and the influence of this on the estimates. The coefficient of $CF_{jt}$ in (4) is significant and the adjusted R-square improves when $CF_{jt}$ is added. While insignificant, the price coefficient of (4) is larger in absolute value than that of (3). These results weakly support the existence of endogeneity in (3). Along with the endogeneity, (5) additionally considers the preference changes (in an admittedly crude manner). I previously mentioned that the price endogeneity may result from time-varying preferences. If (5) handles preference change acceptably, then the endogeneity problem due to time-varying preferences can be alleviated. The estimation result of (5) implies that this is the case. While the estimates of $\Delta\alpha$ and $\Delta\bar{\xi}$ are not significant, the estimates of $\Delta\beta$ are all significant, indicating that the consumer preferences regarding cigarette attributes are actually changing over time. Preferences for "slim", "super-slim", "menthol", and "low-tar" increase over time, whereas those for "long" and "high-tar" decrease over time. I observe a substantial improvement in model fit in (5). Also, the coefficient of $CF_{jt}$ becomes insignificant in (5). All these results imply that an endogeneity problem may occur due to time-varying preferences but this problem can be corrected by considering time-varying preferences appropriately.

**Table 2.1: Estimation Results of (3)-(5)**

| Parameter | (3) | | | (4) | | | (5) | | |
|---|---|---|---|---|---|---|---|---|---|
| | Estimate | SE | t-value | Estimate | SE | t-value | Estimate | SE | t-value |
| $\bar{\xi}$ | -2.23 | 0.13 | -16.6 | -2.08 | 0.14 | -14.8 | -2.55 | 0.21 | -12.2 |
| $\Delta\bar{\xi}$ | | | | | | | 0.20 | 0.25 | 0.8 |
| $\alpha$ | -1.16 | 0.09 | -13.6 | -1.21 | 0.09 | -14.0 | -0.82 | 0.17 | -4.8 |
| $\Delta\alpha$ | | | | | | | -0.25 | 0.19 | -1.3 |
| CF | | | | 0.16 | 0.05 | 3.5 | -0.02 | 0.05 | -0.4 |
| $\beta$ Slim | 0.32 | 0.09 | 3.6 | 0.31 | 0.09 | 3.4 | -0.74 | 0.13 | -5.8 |
| Super-slim | 1.29 | 0.14 | 9.4 | 1.41 | 0.14 | 10.0 | -0.54 | 0.24 | -2.3 |
| Menthol | -1.47 | 0.08 | -18.3 | -1.45 | 0.08 | -18.0 | -1.78 | 0.11 | -15.8 |
| Long | -0.71 | 0.09 | -7.7 | -0.72 | 0.09 | -7.8 | -0.19 | 0.13 | -1.5 |
| Low-tar | -0.55 | 0.09 | -6.0 | -0.52 | 0.09 | -5.7 | -1.03 | 0.14 | -7.2 |
| High-tar | -1.20 | 0.09 | -14.2 | -1.21 | 0.09 | -14.2 | -0.92 | 0.11 | -8.6 |
| $\Delta\beta$ Slim | | | | | | | 1.83 | 0.17 | 10.9 |
| Super-slim | | | | | | | 2.90 | 0.28 | 10.2 |
| Menthol | | | | | | | 0.64 | 0.15 | 4.2 |
| Long | | | | | | | -1.05 | 0.16 | -6.5 |
| Low-tar | | | | | | | 0.73 | 0.18 | 4.0 |
| High-tar | | | | | | | -0.77 | 0.13 | -5.8 |
| P1000 | 0.27 | 0.07 | 3.9 | 0.28 | 0.07 | 4.0 | 0.35 | 0.07 | 4.8 |
| P1500 | 0.07 | 0.10 | 0.7 | 0.04 | 0.10 | 0.4 | 0.08 | 0.10 | 0.8 |
| P2000 | 0.61 | 0.14 | 4.4 | 0.48 | 0.14 | 3.3 | 0.28 | 0.15 | 1.9 |
| PIA1 | -0.33 | 0.13 | -2.5 | -0.34 | 0.13 | -2.5 | -0.36 | 0.13 | -2.9 |
| PIA2 | -0.10 | 0.13 | -0.7 | -0.10 | 0.13 | -0.8 | -0.13 | 0.12 | -1.0 |
| PIB1 | 0.49 | 0.13 | 3.8 | 0.50 | 0.13 | 3.9 | 0.50 | 0.12 | 4.1 |
| PIB2 | 0.09 | 0.13 | 0.7 | 0.10 | 0.13 | 0.8 | 0.10 | 0.12 | 0.9 |
| Jan | -0.20 | 0.09 | -2.2 | -0.20 | 0.09 | -2.2 | -0.20 | 0.09 | -2.3 |
| Feb | -0.28 | 0.09 | -3.0 | -0.28 | 0.09 | -3.0 | -0.27 | 0.09 | -3.1 |
| Age/100 | -0.05 | 0.15 | -0.3 | -0.24 | 0.16 | -1.5 | -0.13 | 0.16 | -0.8 |
| (Age/100)^2 | -0.11 | 0.06 | -1.9 | -0.05 | 0.06 | -0.8 | -0.01 | 0.06 | -0.1 |
| 1/Age | -0.72 | 0.27 | -2.7 | -0.74 | 0.27 | -2.8 | -0.85 | 0.25 | -3.3 |
| R-sq. | | 0.34 | | | 0.35 | | | 0.42 | |
| Adj. R-sq. | | 0.34 | | | 0.34 | | | 0.41 | |
| SE. of Reg. | | 1.09 | | | 1.08 | | | 1.02 | |

I now discuss the estimates of $\rho$. All models lead to similar estimates for $\rho$. The coefficients for $P_{1000}$ and $P_{2000}$ are significantly positive and this supports my hypothesis that consumers prefer round prices, perhaps because of transactional convenience. *PIB1* is significantly positive implying significant stockpiling behavior in the month before the price increase. Also, *PIA1* is significantly negative implying a significant sales decrease in the month of price increase due to the price change, and possibly the earlier stockpiling. Dummies for January and February are significant and this indicates a significant decrease in demand during these months. The estimated coefficients for $Age_{jt}$, $Age^2_{jt}$, and $1/Age_{jt}$ indicate similar trends in the three models. The demand of the newly introduced product is low during the first few months but increases subsequently and plateaus. This is consistent with a process of new product diffusion.
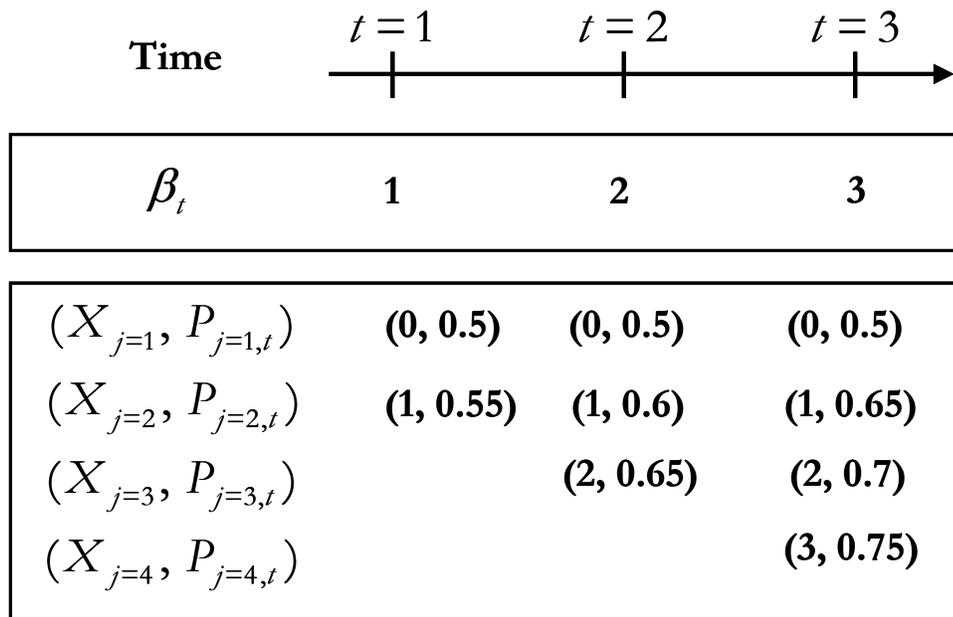
To summarize, in this section, I have shown using a homogeneous logit model that consumer preferences towards product attributes vary over time. Moreover, when ignored, this time-varying pattern causes an econometric problem of price endogeneity under the strategic pricing behavior of the firm.

## 2.3.3 Endogenous New Product Introduction and Attribute Choice

What is the expected action of the firm if consumer preference changes? It may modify its product line by introducing new products which are designed to satisfy consumers' changing preferences. In this section, I study the econometric issue of such strategic behaviors of the firm. I first illustrate the endogenous new product introduction and attribute decision under time-varying consumer preferences using a simple example. Let us assume that we are interested in estimating consumers' preferences with respect to product attributes and their price sensitivity in a differentiated goods market. We observe market shares (or sales) of products and their attributes. After appropriately handling the heterogeneity of consumer

preferences, one can obtain an equation which specifies the relationship among the common utilities, product attributes, prices and consumer preferences:

$\delta_{jt} = X_j' \beta_t + P_{jt} \alpha + \xi_{jt}$. I further assume that there is a single important attribute in this product category and that consumers' mean preference with respect to this attribute is changing over time. In particular, $\beta_{t=1} = 1$, $\beta_{t=2} = 2$, and $\beta_{t=3} = 3$. At $t = 1$, two products ($j = 1$ and $j = 2$) exist in the market with levels of the attribute $X_{j=1} = 0$ and $X_{j=2} = 1$. Also, $P_{j=1,t=1} = 0.5$ and $P_{j=2,t=1} = 0.55$. At $t = 2$, consumer preference changes from $\beta_{t=1} = 1$ to $\beta_{t=2} = 2$. That is, consumer preference toward the attribute is strengthened. Moreover, knowing this preference change, a firm introduces a new product which is fortified in the attribute ($X_{j=3} = 2$). Also, the firm charges increased prices for products with the attribute: $P_{j=2,t=2} = 0.6$ and $P_{j=3,t=2} = 0.65$. Here, I assume that such firm actions are optimal for the firm. At $t = 3$, the trend of preference change continues and another new product is introduced ($X_{j=4} = 3$). Again, the firm charges increased prices for the products with the attribute: $P_{j=2,t=3} = 0.65$, $P_{j=3,t=3} = 0.70$, $P_{j=4,t=3} = 0.75$. In this example, the firm behaves optimally using strategic tools - introduction of new product, product attribute choice, and pricing. Consequently, I observe correlations among the time-varying preferences, prices, and the distributions of product attributes and these may cause an endogeneity problem. Figure 2.3 demonstrates the evolution of market and the correlation among $\beta_t$, $P_{jt}$, and $X_j$.

| Time | $t=1$ | $t=2$ | $t=3$ |
|---|---|---|---|
| $\beta_t$ | 1 | 2 | 3 |

| | $t=1$ | $t=2$ | $t=3$ |
|---|---|---|---|
| $(X_{j=1}, P_{j=1,t})$ | (0, 0.5) | (0, 0.5) | (0, 0.5) |
| $(X_{j=2}, P_{j=2,t})$ | (1, 0.55) | (1, 0.6) | (1, 0.65) |
| $(X_{j=3}, P_{j=3,t})$ | | (2, 0.65) | (2, 0.7) |
| $(X_{j=4}, P_{j=4,t})$ | | | (3, 0.75) |

**Figure 2.3: Time-Varying Preference and Endogenous New Product Introduction**

**And Attribute Choice**

A natural question here is the impact of the correlations on the estimation of demand parameters. Consider the case where a researcher ignores the dynamics in parameters, knowingly or unknowingly, and estimates a static parameter $\beta_{static}$ instead of $\beta_t$. It is clear that $\beta_{static}$ cannot inform us of the exact time path of preferences. The more practical questions here are how representative $\beta_{static}$ is and whether ignoring the dynamics in parameters biases the other parameters. If one is to summarize $\beta_t$ using a single representative value, a natural answer might be its mean value over time ($\beta_{mean}$). Note that $\beta_{mean}$ is equal to 2 ($=\Sigma_t\beta_t/T$) in our example. However, if one ignores the correlation between $\beta_t$ and $\{X_j\}_t$, $\beta_{static}$ will be different from $\beta_{mean}$. If there exists positive (negative) correlation between $\beta_t$ and $\{X_j\}_t$, $\beta_{static}$ will be greater (smaller) than $\beta_{mean}$. Another problem in estimation occurs due to the correlation between $\{X_j\}_t$ and $\beta_t$'s in the new error term. This correlation results from the endogenous new product introduction and attribute choice.[27] As a result, the model suffers from another endogeneity issue as well as the price endogeneity mentioned in section 2.3.2.

To empirically investigate the estimation problems, I perform a simple Monte Carlo experiment. The data generating process is as in Figure 2.3 and $\xi_{jt} \sim iid.N(0,0.1^2)$. The mean and standard deviation of estimates of $\beta_{static}$ over 10,000 repetitions are 2.86 and 0.04, respectively. The estimates are far from $\beta_{mean}$ and provide a biased summary of $\beta_t$'s. The mean and standard deviation of estimates of $\alpha$ are -1.62 and 0.10, respectively. The estimates are significantly different from the true value -1. Note that the direction of the bias is the opposite of what the price endogeneity due to the correlation between $\beta_t$ and $P_{jt}$ predicts. Since $\beta_t$ and $P_{jt}$ are

---

[27] Note that this econometric issue due to the correlation between $\beta_t$ and $\{X_j\}_t$ is similar to the "slope endogeneity" problem (Villas-Boas and Winer 1995; Kuksov and Villas-Boas 2008; Luan and Sudhir 2009). The major difference is that the correlation between $\beta_t$ and $\{X_j\}_t$ is a result of permanent shocks, while the "slope endogeneity" problem is a result of temporary shocks.

positively correlated, I should observe an estimate greater than -1. This is attributed to endogeneity due to the correlation between $\{X_j\}_t$ and $\beta_t$'s. Note that the direction of bias is similar to what we have observed in Section 2.3.2. The estimated price coefficient of the static model (3) is -1.164, while that of the two-period model (5) is -0.820.

Now consider the case where the preferences are estimated piecewise over time. That is, $\beta_t$ is estimated as it is in $\delta_{jt} = X'_j \beta_t + P_{jt} \alpha + \xi_{jt}$. Note that there is no correlation between $\beta_t$ and $X_j$ at a given time $t$. Thus, the estimation is not subject to endogeneity due to the correlation between $\beta_t$'s and $X$. Also note that the price endogeneity coming from the use of $\beta_{static}$ instead of $\beta_t$ is no longer a problem. Consequently, we can easily get consistent estimates of parameters using standard estimators. Note that the time-varying parameter model belongs to this case and therefore, is free from endogeneity issues. I estimate a model with dummy variables to handle parameter dynamics:

$\delta_{jt} = X_j \beta_1 + X_j \cdot \Delta\beta_2 \cdot I(t=2) + X_j \cdot \Delta\beta_3 \cdot I(t=3) + P_{jt} \alpha + \xi_{jt}$. The mean (standard deviation) of estimates of $\beta_1$, $\Delta\beta_2$, $\Delta\beta_3$, and $\alpha$ over 10,000 repetitions are 1.00 (0.11), 1.00 (0.11), 2.00 (0.10), and -1.00 (0.10), respectively. The true values of all parameters are perfectly recovered. To summarize implications from this section, when consumers' preferences are changing over time, time-varying parameter models should be employed because they provide detailed information on the time path of preferences and circumvent problems due to endogenous price and other regressors.

## 2.4 Model and Estimation

I propose a structural model which will be applied to the data to analyze both demand and supply sides of the South Korean cigarette market. The model consists of three sub-models; 1) demand model, 2) pricing model, and 3) new product introduction and attribute choice model. I use the implications of the preliminary

analyses in building these models. After proposing the model, I provide the details on the estimation of the proposed model.

## 2.4.1 Demand Model

The main additional features of the demand model compared to models in the preliminary analysis are parameter dynamics and heterogeneity in consumer preference. At each time period $t = 1,...,T$, the utility of brand $j=0,...,J_t$ for consumer $h=1,...,H$ is given by the following expressions:

$$U_{h0t} = \tilde{\xi}_{0t} + \varepsilon_{h0t}, \qquad (6)$$

$$U_{hjt} = X'_j \beta_{ht} + P_{jt} \alpha_h + Y'_{jt} \rho + \bar{\xi} + CF_{jt} \vartheta + \hat{\xi}_{jt} + \varepsilon_{hjt}, \quad j=1,...,J_t, \qquad (7)$$

$$\alpha_h = \alpha + a_h, \qquad (8)$$

$$\beta_{ht} = \beta_t + b_h, \qquad (9)$$

where an (1×1) scalar $a_h$ follows $N(0,\sigma_a^2)$, a (6×1) vector $b_h$ follows $N(0,\Sigma_b)$, $\Sigma_b$ is a diagonal matrix, and a (6×1) vector $\beta_t = [\beta_t^{sl} \ \beta_t^{ss} \ \beta_t^{mt} \ \beta_t^{lg} \ \beta_t^{lt} \ \beta_t^{ht}]'$. Note that this model incorporates heterogeneity in consumers' preferences for cigarette attributes and price sensitivity as specified in (8) and (9). To capture dynamics in consumer preferences, I specify the following evolution of $\beta_t^g$ for $g \in \{sl, ss, mt, lg, lt, ht\}$,

$$\beta_t^g = \beta_{t-1}^g + \varphi^g \eta_{t-1}^g + \varsigma_t^g, \qquad (10)$$

where $\eta_{t-1}^g = I(\text{a product with attribute level "g" is introduced to the market during t-1})$, $\varphi^g$ is a (1×1) scalar coefficient, and a (1×1) scalar $\varsigma_t^g$ follows $N(0,\sigma_{\varsigma,g}^2)$. $\varphi^g \eta_{t-1}^g$ and $\varsigma_t^g$ capture the introduction effect and the exogenous effect, respectively. (10) implies that the current preference is constructed by adding the exogenous effect and the introduction effect to the preference of previous period. The dynamics of the exogenous effect follows a random walk process since it is highly versatile in modeling various time trends and structural breaks. To handle time-varying parameters in the estimation, I incorporate the Kalman Filtering algorithm into the

likelihood-based aggregate RCL estimation method. We provide more details on the estimation in Section 2.4.4.

One may question the advantages of the proposed demand model over the structural break model (e.g. (5)). The structural break model can approximate continuous changes in $\beta_t$ at best. Moreover, researchers must have good knowledge of the number of breaks and the timing of each break, otherwise the result can be misleading. However, researchers do not have such knowledge in many cases. In contrast, the proposed model does not require such knowledge. The structural break model divides the sample period into several sub-periods. In each sub-period, the structural break model is nothing but a static model. Therefore, the structural model can suffer from the same problems as the static model within each sub-period unless the number of sub-periods is not large. All in all, the proposed model is complete in the handling of time-varying parameters.

## 2.4.2 Pricing Model

In this section I develop the firm's pricing model. To this end, two important facets of the market are noteworthy. First, as detailed previously, cigarette prices are monitored and influenced by the government. This implies a strong possibility that the price may not be set at the profit-maximizing level. In marketing and industrial organization literatures, price models based on an assumption of profit maximization have been popular (e.g. BLP 1995; Sudhir 2001). However, such a pricing model is likely to be inapplicable to cigarette prices in South Korea. Second, the unique feature of demand is the evolution of consumer preferences for cigarette attributes as is confirmed in the preliminary analysis using a homogenous logit model. Accordingly the proposed demand model incorporates this using the time-varying parameter specification of (10). An interesting question that arises here is what is the implication of time-varying consumer preferences on the firm's pricing decision. Faced with

changing preferences for product attributes, the firm might take advantage of this and strategically change prices.

I propose the following price model which appropriately considers above-mentioned key aspects of the market:

$$P_{jt} = \gamma_c + t \cdot \gamma_t + X'_j \gamma_x + (X_j^{sl} \beta_t^{sl}) \gamma_{sl} + (X_j^{ss} \beta_t^{ss}) \gamma_{ss} +$$
$$(X_j^{mt} \beta_t^{mt}) \gamma_{mt} + (X_j^{lg} \beta_t^{lg}) \gamma_{lg} + (X_j^{lt} \beta_t^{lt}) \gamma_{lt} + (X_j^{ht} \beta_t^{ht}) \gamma_{ht} + \varepsilon_{jt}^P \qquad , j = 1,...,J_t \quad (11)$$

where $\gamma_c$, $\gamma_t$, $\gamma_{sl}$, $\gamma_{ss}$, $\gamma_{mt}$, $\gamma_{lg}$, $\gamma_{lt}$, and $\gamma_{ht}$ are (1×1) scalars and $\gamma_x = [\ \gamma_{x,sl}\ \ \gamma_{x,ss}$ $\gamma_{x,mt}\ \ \gamma_{x,lg}\ \ \gamma_{x,lt}\ \ \gamma_{x,ht}]'$ is a (6×1) vector. This model is a linear regression of prices on a constant, a time trend, product attributes, and interactions of time-varying parameters and product attributes. It is clear that this model is free from any assumption on the firm behavior. This model specification reduces the risk of misspecification which might result in misleading implications about pricing behavior as well as introduce bias into the new product introduction and attribute choice model, which is discussed subsequently.

Time trend $t$ is to capture increasing trend in cigarette prices. Dummy variables for product attributes $X_j$ enter into the model to capture their influence on the prices of products possessing those attributes. This specification is similar to a hedonic price model which decomposes a product into its constituent characteristics, and obtains estimates of the contributory value of each characteristic to the product's price. The interactions of time-varying parameters and product attributes are included to capture the relationship between the change in preference of each attribute and the change in prices of the relevant products. As an illustration, assume that $\gamma_{x,sl}$ is positive. This implies that when the preference toward the attribute "slim" increases, the relative prices of products with this attribute also increase. $\varepsilon_{jt}^P$ represents the

remaining unexplained part in prices. I assume that $\varepsilon_{jt}^{P}$ follows a normal distribution with mean zero and variance $\sigma_{P}^{2}$.

### 2.4.3 New Product Introduction and Attribute Choice Model

In this section, I propose the new product introduction and attribute choice model which explains the firm's decision process regarding the new product. As mentioned previously, the strategic manipulation of price, advertising, and promotion is strongly restricted in this market. Therefore, the most important marketing activity of the firm is designing and introducing a new product. In empirical marketing and industrial organization literature, the firm's decision on new product introduction and its design or attribute decision has not been frequently studied compared to short-term strategic decisions (i.e. price, advertising, and promotion). This can be partially attributed to the data availability. Researchers can commonly observe variations in firms' short-run strategic decisions over time and/or across markets. Thus, it is relatively easy to obtain data suitable to study firms' decisions on price, advertising, and promotion. However, when it comes to long-run decisions, such as firm's decisions regarding new product introduction and its design, there is not much variation in these variables in most datasets. This is why the long-run decisions have been routinely maintained as exogenous (Dubé et al. 2005). In this study, the consumer preference change provides rich information that help identify the firm behavior. As consumer preferences evolve over time, the optimal new product design and its profitability vary as well. This provides informative variations that help identify the firm's decision process regarding whether to introduce a new product, and its optimal design. This is a unique feature of my analysis compared to extant studies that analyze firms' new product introduction or product positioning decisions.

At the beginning of time period $t$, $J_t$ products are available to consumers in the market. These products are the same as were offered at the end of $t\text{-}1$. I refer to them

as existing products and denote them as $EP_t$. During time period $t$, a new product can be introduced from a set of fifteen product types. This set is a subset of the feasible combinations of four attributes (i.e. thickness, length, menthol taste, and tar content). Three levels of thickness (regular, slim, or super-slim), two levels of length (standard type or long type), two levels of menthol taste (yes or no), and three levels of tar content (high, medium, and low) result in 36 (=3×2×2×3) different product profiles. However, I exclude three specific pairings of attribute levels: 1) menthol and long; 2) slim and 84 mm; 3) super-slim and 84 mm. These exclusions are based on the observation that these combinations are almost never observed in the market, not only during the sample period and but also before and after. Consequently, the set of candidate products (*CP*) shrinks to fifteen profiles or product types, and I denote them as $k \square CP = \{1,...,15\}$. If $k$ is introduced into the market during $t$, $EP_{t+1}$ becomes $EP_t \cup \{k\}$. Table 2.2 shows details on attribute levels of candidate product types and their observed introduction frequencies during the sample period.

I assume that at most one product can be newly introduced into the market during each time period. This assumption is reasonable in my data where no more than one introduction occurs in any month. Also, this assumption helps me frame the problem of new product introduction and attribute choice in the well-known discrete choice setting. Along with 17 new product introductions, I observe that the firm withdraws four products from the market in the data. However, for simplicity I do not model the product withdrawal decision and regard the changes in $EP_t$ due to product exits as exogenous.[28]

---

[28] During the 106 month sample period, I observe only four drops. According to KT&G personnel, the drop decision is not as critical as the introduction decision since keeping the production of existing products does not impose substantial cost to the company given the size of current product line and its manufacturing capacity. This explains why I observe such a small number of drops.

**Table 2.2: Candidate Product Profiles and Introduction Frequencies**

| $k$ | Thickness | Menthol | Length (mm.) | Tar | # of introductions |
|---|---|---|---|---|---|
| 1 | Regular | No | 84 | High | 1 |
| 2 | Regular | No | 84 | Medium | 6 |
| 3 | Regular | No | 84 | Low | 3 |
| 4 | Regular | No | $\geq 100$ | High | 0 |
| 5 | Regular | No | $\geq 100$ | Medium | 1 |
| 6 | Regular | No | $\geq 100$ | Low | 0 |
| 7 | Regular | Yes | 84 | High | 0 |
| 8 | Regular | Yes | 84 | Medium | 1 |
| 9 | Regular | Yes | 84 | Low | 0 |
| 10 | Slim | No | $\geq 100$ | High | 0 |
| 11 | Slim | No | $\geq 100$ | Medium | 2 |
| 12 | Slim | No | $\geq 100$ | Low | 0 |
| 13 | Super-slim | No | $\geq 100$ | High | 0 |
| 14 | Super-slim | No | $\geq 100$ | Medium | 1 |
| 15 | Super-slim | No | $\geq 100$ | Low | 2 |

I specify the expected utilities and expected prices of candidate product

$k \square \ CP = \{1,...,15\}$ at $t$ as follow:

$$
\begin{aligned}
U_{hkt} &= (X'_k \beta_t + P_{kt}(\beta_t, X_k)\alpha + Y'_{kt}\rho) + (X'_k b_h + P_{kt}(\beta_t, X_k)a_h) + \varepsilon_{hkt}, \quad (12)\\
&= \delta_{kt} + \mu_{hkt} + \varepsilon_{hkt}
\end{aligned}
$$

$$
\begin{aligned}
P_{kt}(\beta_t, X_k) &= \gamma_0 + t \cdot \gamma_t + X'_k \gamma_x + (X^{sl}_k \beta^{sl}_t)\gamma_{sl} + (X^{ss}_k \beta^{ss}_t)\gamma_{ss} + \\
&\quad (X^{mt}_k \beta^{mt}_t)\gamma_{mt} + (X^{lg}_k \beta^{lg}_t)\gamma_{lg} + (X^{lt}_k \beta^{lt}_t)\gamma_{lt} + (X^{ht}_k \beta^{ht}_t)\gamma_{ht}
\end{aligned}
\quad (13)
$$

where $\delta_{kt} = (X'_k \beta_t + P_{kt}(\beta_t, X_k)\alpha + Y'_{kt}\rho)$, $\mu_{hkt} = (X'_k b_h + P_{kt}(\beta_t, X_k)a_h)$, and $\varepsilon_{hkt}$ is the unexplained part of utility which is assumed to follow type I extreme value distribution. Here I assume that the firm's price decision on the new product is in line with (11). Note that (13) is the expected value of $P_{kt}$ given $\beta_t$ and $X_k$ in the light of (11) and this is plugged into (12). This is consistent with a two-step decision process of the firm where the product type is decided first and then price is decided based on product attributes $X_k$ and consumer preferences for these attributes at $t$, $\beta_t$. From (12) and (13), I can specify the market share of product $j$ at $t$ assuming the new introduction of $k$, or no new introduction.

$$
S_{jt}(0) = \int \frac{exp(\delta_{jt} + \mu_{hjt})}{1 + \Sigma^{J_t}_{i=1} \exp(\delta_{it} + \mu_{hit})} f(a_h, b_h) d(a_h, b_h), \quad (14)
$$

$$
S_{jt}(k) = \int \frac{exp(\delta_{jt} + \mu_{hjt})}{1 + \Sigma^{J_t}_{i=1} \exp(\delta_{it} + \mu_{hit}) + \exp(\delta_{kt} + \mu_{hkt})} f(a_h, b_h) d(a_h, b_h), \quad (15)
$$

where $S_{jt}(0)$ denotes the market share of $j = 1,..., J_t$ at $t$ when no product is introduced during that period and $S_{jt}(k)$ denotes the market share of $j = 1,..., J_t, k$ at $t$ when $k \square \ CP$ is introduced during that period. The profit of the firm can be calculated as follows:

$$
\pi_t(0) = \Sigma^{J_t}_{j=1}\left(M_t \times S_{jt}(0) \times (W_{jt} - MC_{jt} - Tax_{jt})\right), \quad (16)
$$

$$\pi_t(k) = \Sigma_{j=1}^{J_t}\Big(M_t \times S_{jt}(k) \times (W_{jt} - MC_{jt} - Tax_{jt})\Big) \\ + M_t \times S_{kt}(k) \times (W_{kt} - MC_{kt} - Tax_{kt}) - F \quad, \quad k \square \ CP, \tag{17}$$

where $\pi_t(0)$ denotes the profit at $t$ when no product is introduced during that period, and $\pi_t(k)$ denotes the profit at $t$ when $k$ is introduced during that period. $\{W_{jt}\}_{j=1,...,J_t,k}$ denotes wholesale prices and $M_t$ denotes the market potential at $t$. $MC_{jt}$, $Tax_{jt}$, and $F$ denote marginal cost of producing $j$ at $t$, the tax imposed on $j$ at $t$, and the fixed cost of adding one product to the existing product line, respectively. Using (16) and (17), I can specify the firm's utility of introducing $k$ ($V_t(k)$) and that of no introduction ($V_t(0)$) at $t$ as follows:

$$V_t(0) = \pi_t(0) + e_{0t}, \tag{18}$$

$$V_t(k) = \pi_t(k) + e_{kt}, \quad k \square \ CP = \{1,...,15\}. \tag{19}$$

The explained part in the firm's utility is the profit from the sales of products. The remaining unexplained part of the utility is represented by $e_{0t}$ and $\{e_{kt}\}_{k \square CP}$ which follow type I extreme value distributions with variance $\sigma_e^2$. The firm will introduce $k$ if $V_t(k) > V_t(l)$ for $l \neq k$, $l \square \ \{0,1,...,15\}$. Note that by normalizing $V_t(0)$ and $\{V_t(k)\}_{k \square CP}$, I can derive the logit probabilities for the firm to introduce $k$ ($PI_t^M(k)$) and not to introduce any product ($PI_t^M(0)$) at $t$:

$$PI_t^M(0) = \frac{exp(\widetilde{\pi}_t(0))}{\Sigma_{i=0}^{15} \exp(\widetilde{\pi}_t(i))}, \tag{20}$$

$$PI_t^M(k) = \frac{exp(\widetilde{\pi}_t(k))}{\Sigma_{i=0}^{15} \exp(\widetilde{\pi}_t(i))}, \quad k \square \ CP, \tag{21}$$

where $\widetilde{\pi}_t(0) = \pi_t(0) \times \pi/(\sigma_e \sqrt{6})$ and $\widetilde{\pi}_t(k) = \pi_t(k) \times \pi/(\sigma_e \sqrt{6})$.

**2.4.4 Estimation**

The specification of the demand model in (6)-(9) is similar to that of BLP (1995) or Nevo (2001) in that the utility is specified at the individual level with consumer heterogeneity while the aggregate data are used in the estimation, and in that

the utility contains the common demand shocks which are correlated with the price. A distinctive feature of the proposed demand model is that the coefficients for the product attribute $\beta_t$ are stochastically evolving over time as specified in (10). To handle the heterogeneity, I apply the method proposed by BLP (1995). To be more specific, I separate the common part and the consumer-specific deviation in the utility using the BLP contraction mapping. I correct for price endogeneity using the control function method (Petrin and Train 2009; Park and Gupta 2009). For the inference of unobservable time-varying coefficients, I use the Kalman filtering algorithm.

After assuming that $\tilde{\xi}_{0t} = 0$ for normalizing, I can rewrite (6)-(7) as follows:

$$U_{hjt} = X'_{hj}\beta_{ht} + P_{jt}\alpha_h + Y'_{jt}\rho + \overline{\xi} + CF_{jt}\vartheta + \hat{\xi}_{jt} + \varepsilon_{hjt} \tag{6'}$$

$$= (X'_j\beta_t + P_{jt}\alpha + Y'_{jt}\rho + \overline{\xi} + CF_{jt}\vartheta + \hat{\xi}_{jt}) + (X'_jb_h + P_{jt}a_h) + \varepsilon_{hjt} . \tag{7'}$$

Since $\varepsilon_{h0t}$ and $\delta_{\{\varepsilon_{hjt} + \mu_{hjt} + \varepsilon_{hjt}\}_{j=1,...,J_t}}$ are from type-I extreme value distribution, (6')-(7') imply the following choice probabilities given $\sigma_a^2$ and $\Sigma_b$:

$$S_{jt}(\delta_{jt} \mid \sigma_a^2, \Sigma_b) = \int \frac{exp(\delta_{jt} + \mu_{hjt})}{1 + \Sigma_{i=1}^{J_t} \exp(\delta_{it} + \mu_{hit})} \phi(a_h, b_h \mid \sigma_a^2, \Sigma_b) d(a_h, b_h). \tag{22}$$

For the model estimation, I first assume that the observed market share $s_{jt}$ is equal to $S_{jt}(\delta_{jt} \mid \sigma_a^2, \Sigma_b)$ or $s_{jt} = S_{jt}(\delta_{jt} \mid \sigma_a^2, \Sigma_b)$. Note that this assumption is different from that of a usual discrete choice model in which the observation is the multinomial outcome from choice probabilities. However, this assumption is reasonable when the $H$ is large enough as in BLP (1995) and my case. Berry (1994) proves that the function $S_{jt}(\cdot \mid \sigma_a^2, \Sigma_b)$ is invertible and any observed vector of market shares can be explained by a unique vector of $\delta_{jt}$ given $\sigma_a^2$ and $\Sigma_b$. That is, $\delta_{jt}(\sigma_a^2, \Sigma_b) = S_{jt}^{-1}(s_{jt} \mid \sigma_a^2, \Sigma_b)$ with abuse of notation. I can compute $\delta_{jt}(\sigma_a^2, \Sigma_b)$ using the contraction mapping proposed in BLP (1995). By replacing $\delta_{jt}$ with $\delta_{jt}(\sigma_a^2, \Sigma_b)$, I have the following equation:

$$\delta_{jt}(\sigma_a^2, \Sigma_b) = X'_j\beta_t + P_{jt}\alpha + Y'_{jt}\rho + \overline{\xi} + CF_{jt}\vartheta + \hat{\xi}_{jt}. \tag{23}$$

75

(23) is a linear time-varying parameter model with an endogenous explanatory variable. To overcome potential problems due to endogeneity, the control function $CF_{jt}$ is introduced into the model. $CF_{jt}$ is the residual from the regression of $P_{jt}$ on the IV's.[29] In the estimation, I jointly estimate the regression model to create $CF_{jt}$. I denote the log-likelihood function of this model as $l_{CF}(\Theta_{CF})$.

The remaining issue in the estimation of demand model is parameter dynamics. As specified in (10), $\beta_t^g$ follows an extended Gaussian random walk process $\beta_t^g = \beta_{t-1}^g + \varphi^g \eta_{t-1}^g + \varsigma_t^g$, where $\varsigma_t^g \sim N(0, \sigma_{\varsigma,g}^2)$. Assuming that $\hat{\xi}_{jt}$ follows $N(0, \sigma_\xi^2)$, I can derive the following Kalman filtering equations:

$$\beta_{t|t-1} = \beta_{t-1|t-1} + \Phi \eta_{t-1},$$

$$V_{t|t-1} = V_{t-1|t-1} + \Sigma_\varsigma,$$

$$R_{t|t-1} = \delta_t(\sigma_a^2, \Sigma_b) - X_t \beta_{t|t-1} - P_t \alpha - Y_t \rho - \bar{\xi} \iota_{J_t} - CF_t \vartheta,$$

$$F_{t|t-1} = X_t V_{t|t-1} X_t' + \sigma_\xi^2 \cdot I_{J_t},$$

$$\beta_{t|t} = \beta_{t|t-1} + V_{t|t-1} X_t' F_{t|t-1}^{-1} R_{t|t-1},$$

$$V_{t|t} = V_{t|t-1} - V_{t|t-1} X_t' F_{t|t-1}^{-1} X_t V_{t|t-1}',$$

where $\psi_t$ denotes a set of information up to $t$, $\Phi$ is a diagonal matrix of $\varphi^g$'s, $\eta_{t-1}$ is a vector of $\eta_{t-1}^g$'s, $\beta_{t|t-1} = E[\beta_t | \psi_{t-1}]$, $\beta_{t|t} = E[\beta_t | \psi_t]$,

$V_{t|t-1} = E[(\beta_t - \beta_{t|t-1})(\beta_t - \beta_{t|t-1})']$, $V_{t|t} = E[(\beta_t - \beta_{t|t})(\beta_t - \beta_{t|t})']$, $F_{t|t-1} = E[R_{t|t-1}^2]$,

$R_{t|t-1} = \delta_t(\sigma_a^2, \Sigma_b) - E[\delta_t(\sigma_a^2, \Sigma_b) | \psi_{t-1}]$, $\delta_t(\sigma_a^2, \Sigma_b) = [\delta_{1t}(\sigma_a^2, \Sigma_b) \cdots \delta_{J_t,t}(\sigma_a^2, \Sigma_b)]'$,

$P_t = [P_{1t} \cdots P_{J_t,t}]'$, $CF_t = [CF_{1t} \cdots CF_{J_t,t}]'$, $\iota_{J_t}$ is a ($J_t \times 1$) vector of ones, $X_t$ is a ($J_t \times 6$) matrix with row vectors $\{X_j'\}_{j=1,\dots,J_t}$ and $Y_t$ is a ($J_t \times 12$) matrix with row vectors $\{Y_{jt}'\}_{j=1,\dots,J_t}$. Note that these equations can be recursively applied once the starting

---

[29] In linear models such as (24), the IV method is the most popular approach for handling the endogeneity problem. In this method, one regresses the endogenous variables on instrumental variables and then uses the fitted values instead of the endogenous regressors. However, this method fails in the presence of time-varying parameters (i.e. Markov regime-switching models and state-space models). In these cases, one should use the control function approach. See Kim (2004; 2006) for more details.

values $\beta_{0|0}$ and $V_{0|0}$ are given.[30] I can estimate the model parameters $\Theta_{Demand} = \{\alpha, \rho, \overline{\xi}, \vartheta, \sigma_a^2, \sigma_\xi^2, \Sigma_\varsigma, \Sigma_b, \Phi\}$ by maximizing the following log-likelihood function:

$$
\begin{aligned}
l_{Demand}(\Theta_{Demand}; \Theta_{CF}) = &-0.5 \sum_{t=1}^{T} \ln((2\pi)^{J_t} \cdot | F_{t|t-1} |) \\
&-0.5 \sum_{t=1}^{T} R'_{t|t-1} F_{t|t-1}^{-1} R_{t|t-1} - \sum_{t=1}^{T} | J_{\tilde{\xi} \to S,t} |
\end{aligned}
\tag{24}
$$

where $J_{\tilde{\xi} \to S,t}$ is the Jacobian matrix corresponding to the transformation of variables of $\hat{\xi}_{jt}$'s to market shares. The log-likelihood function depends on $CF_{jt}$ and thus $\Theta_{CF}$ appears in it. The element in row $m$ and column $n$ of $J_{\tilde{\xi} \to S,t}$, $J_{\tilde{\xi} \to S,t}^{m,n}$ is defined as follows:

$$
J_{\tilde{\xi} \to S,t}^{m,n} =
\begin{cases}
\int \left(1 - \dfrac{\exp(\delta_{mt} + \mu_{bmt})}{1 + \Sigma_{i=1}^{J_t} \exp(\delta_{it} + \mu_{bit})}\right)\left(\dfrac{\exp(\delta_{mt} + \mu_{bmt})}{1 + \Sigma_{i=1}^{J_t} \exp(\delta_{it} + \mu_{bit})}\right) \\
\quad \phi(a_b, b_b \mid \sigma_a^2, \Sigma_b)\, d(a_b, b_b), \quad m = n \\
-\int \left(\dfrac{\exp(\delta_{mt} + \mu_{bmt})}{1 + \Sigma_{i=1}^{J_t} \exp(\delta_{it} + \mu_{bit})}\right)\left(\dfrac{\exp(\delta_{nt} + \mu_{bnt})}{1 + \Sigma_{i=1}^{J_t} \exp(\delta_{it} + \mu_{bit})}\right) \\
\quad \phi(a_b, b_b \mid \sigma_a^2, \Sigma_b)\, d(a_b, b_b), \quad m \neq n
\end{cases}
.
$$

Note that the parameters regarding the dispersion of heterogeneity distribution, $\sigma_a^2$ and $\Sigma_b$, are also estimated by maximizing (24). My estimation approach is in line with the likelihood-based estimator of the RCL model using aggregate data (Jiang et al. 2009). Compared to the Generalized Method of Moment (GMM) approach which was originally suggested, the likelihood approach has several advantages. First, it enables easy extensions to likelihood-based models or methods. In my case, the method to handle unobservable time-varying parameters (i.e. Kalman filter) could easily be incorporated into the RCL model using aggregate data. The second advantage is that

---

[30] I used a (6×1) zero-vector and $10 \cdot I_6$ for $\beta_{0|0}$ and $V_{0|0}$, respectively. I tried different starting values and found that the estimation result is not sensitive to starting values.

it enables us to easily perform structural tests for the nature of the market or the firm behavior by evaluating the likelihood values under different scenarios (i.e. likelihood ratio test, Wald test). In the empirical analysis, along with the proposed model, I estimate a benchmark model in which the firm believes that the consumer preferences are not varying over time and makes decisions under such belief. I perform an empirical test of firm behavior comparing the likelihood values of two models.

The estimation of the pricing model is straight-forward. I can estimate pricing model parameters $\Theta_{Price} = \{\gamma_c, \gamma_t, \gamma_{sl}, \gamma_{ss}, \gamma_{mt}, \gamma_{lg}, \gamma_{lt}, \gamma_{ht}, \gamma_x, \sigma_P^2\}$ by maximizing the following log-likelihood function:

$$l_{Price}(\Theta_{Price}; \Theta_{Demand}, \Theta_{CF}) = -0.5 \cdot \left(\sum_{t=1}^{T} J_t\right) \cdot \ln(2\pi\sigma_p^2) - 0.5 \sum_{t=1}^{T} \sum_{j=1}^{J_t} \left(\left(P_{jt} - \gamma_c - t \cdot \gamma_t - \right.\right.$$
$$X_j' \gamma_x - (X_j^{sl}\beta_{t|t}^{sl})\gamma_{sl} - (X_j^{ss}\beta_{t|t}^{ss})\gamma_{ss} - (X_j^{mt}\beta_{t|t}^{mt})\gamma_{mt} -$$
$$\left.\left.(X_j^{lg}\beta_{t|t}^{lg})\gamma_{lg} - (X_j^{lt}\beta_{t|t}^{lt})\gamma_{lt} - (X_j^{ht}\beta_{t|t}^{ht})\gamma_{ht}\right)^2 / \sigma_p^2\right)$$

Note that the pricing model (11) has $\beta_t$'s as regressors but they are unobservable parameters. In the likelihood function, I use the expected values of $\beta_t$'s ($\beta_{t|t} = E[\beta_t | \psi_t]$) given the demand model parameters. Thus, the likelihood function depends on $\Theta_{Demand}$ and $\Theta_{CF}$ as well as $\Theta_{Price}$.

For the estimation of the new product introduction and attribute choice model, I incorporate information on the wholesale prices $W_{jt}$, the marginal costs $MC_{jt}$, and the taxes $Tax_{jt}$. The retail margin is 10% of the retail price during the sample period. Thus, I let $W_{jt} = 0.9 \cdot P_{jt}$ and $W_{kt} = 0.9 \cdot P_{kt}(\beta_t, X_k)$. During the sample period, taxes were increased in July 1996, January 1999, January 2001, and February 2002. Per pack sold, a fixed amount of taxes were imposed but since January 1999, an additional 10% value added tax was imposed. Based on this information, I calculate $\{Tax_{jt}\}_{j=1,...,J_t, t=1,...,T}$ and $\{Tax_{kt}\}_{k \in CP, t=1,...,T}$. For the marginal cost, I specify the following relationship between marginal cost and product attributes:

$$MC_{jt} = c_0 + X'_j c_x, \quad j = 1,...,J_t,$$

$$MC_{kt} = c_0 + X'_k c_x, \quad k \square CP$$

where $c_0$ is a scalar and $(6 \times 1)$ vector $c_x = [c_{x,sl} \quad c_{x,ss} \quad c_{x,mt} \quad c_{x,lg} \quad c_{x,lt} \quad c_{x,ht}]'$. The specification assumes that the marginal cost of a product is determined by its constituent attributes. Additionally, the following equation regarding the marginal cost is derived based on the information from the annual reports of KT&G:

$$\frac{1}{T}\sum_{t=1}^{T}\left(\sum_{j=1}^{J_t}\frac{MC_{jt} \cdot Sales_{jt}}{\left(\Sigma_{j=1}^{J_t} Sales_{jt}\right)}\right) = 206(KRW).^{31}$$

That is, the share-weighted average of marginal costs across products over time is 206 KRW. This provides useful information on the relationship between $c_0$ and $c_x$. To be more specific, $c_0 = 206 - (1/T) \cdot \Sigma_{t=1}^{T}\left(\Sigma_{j=1}^{J_t}((X'_j c_x) \cdot Sales_{jt})/\Sigma_{j=1}^{J_t}Sales_{jt}\right)$. I use this as a restriction in the estimation. As a result, I obtain the estimate of $c_0$ from $c_x$ without direct estimation. The parameters of the new product introduction and attribute choice model, $\Theta_{Intro} = \{c_x, F, \sigma_V\}$ can be estimated by maximizing the following log-likelihood function:

$$l_{Intro}(\Theta_{Intro};\Theta_{Price},\Theta_{Demand},\Theta_{CF}) = \ln\left(\prod_{t=1}^{T}\prod_{k'=0}^{15}\left(PI_t^M(k')\right)^{Intro_{k't}}\right),$$

where $Intro_{kt} = 1$ if $k$ is introduced at $t$, and $Intro_{0t} = 1$ if none is introduced at $t$.

I can jointly estimate all the parameters by maximizing the following log-likelihood function:

$$l_{CF}(\Theta_{CF}) + l_{Demand}(\Theta_{Demand};\Theta_{CF}) + l_{Price}(\Theta_{Price};\Theta_{Demand},\Theta_{CF}) +$$

$$l_{Intro}(\Theta_{Intro};\Theta_{Price},\Theta_{Demand},\Theta_{CF}).$$

---

[31] I assume that $\Sigma_{j=1}^{J_t}(MC_{jt} \cdot Sales_{jt})/(\Sigma_{j=1}^{J_t}Sales_{jt})$ is equivalent to "cost of goods sold"/"quantity sold". I gathered information on "cost of goods sold" and "quantity sold" from the annual reports of the KT&G.

Or, one can perform the estimation in multiple steps. Although such multi-step

estimation may reduce the efficiency of the estimates, it has its advantages: the

consistency of the demand estimates does not depend on the supply side models and

computational burden is reduced. In the empirical analysis, all the integrals in the

likelihood functions are approximated through Monte Carlo simulation methods and

thus, the resulting estimator is a Simulated Maximum Likelihood Estimator (see

Keane 1993).

**2.5 Result and Discussions**

**2.5.1 Estimation Results of Demand Model**

Along with the proposed demand model, I estimate a benchmark model where

the consumer preferences are not varying over time. That is $\beta_t = \beta_{t-1}$ for all $t$ instead

of (10). Table 2.3 shows the estimation results of both the proposed model and the

benchmark model. Also, it provides the three model fit measures of both models – the

log-likelihood, AIC, and BIC. All three measures favor the proposed model over the

benchmark model.

The estimated price coefficient of the proposed model is -0.821 and that of the

benchmark model is -1.222. The considerable difference between the two price

coefficients is mainly attributed to the endogenous new product introduction and

attribute choice, which the benchmark model does not properly account for. When the

time-varying preferences are ignored, the price coefficient increases by 55%. The

lesson from this observation is clear. If the consumer preferences are changing over

time and the firm considers this in their decision making, time-varying preferences

should be properly modeled. Otherwise, it can spoil not only the preference estimates

but also other important estimates. In my empirical analysis of the cigarette market,

the price coefficient is crucial. Based on this, public policies are determined, as well

as the firm's marketing decisions. For example, if the absolute value of the price

coefficient is large, implying that the consumers are sensitive to cigarette prices, government policies of increasing cigarette prices to decrease the smoking rate can be supported. If the absolute value of the price coefficient is small, the policies using the price measure may lose support. Considering the importance of such issues, proper handling of time-varying preference and subsequent firm behavior should be emphasized.

The coefficient for $CF_{jt}$ is significantly positive in the benchmark model while that of the proposed model is not significantly different from zero. The UPC term in the benchmark model contains the influence of $\beta_t$ as well as $\xi_{jt}$ since the model ignores the time-varying aspect of preferences. Thus, $CF_{jt}$ in the benchmark model controls two different correlations: the correlation between the prices and $\beta_t$ and the correlation between the prices and $\xi_{jt}$. However, $CF_{jt}$ in the proposed model controls only the correlation between the prices and $\xi_{jt}$ since the time-varying preferences $\beta_t$ is appropriately considered in the model.[32] The estimation result implies that the correlation between the prices and $\beta_t$ is significant and positive while the correlation between the prices and $\xi_{jt}$ is not significantly different from zero. The proposed pricing model directly examines the correlation between prices and $\beta_t$. The estimation result of the pricing model also points out the significant positive correlation between the prices and $\beta_t$, echoing what I find here.

---

[32] $CF_{jt}$ is the residual from the regression of $P_{jt}$ on the IV's and estimated by maximizing $l_{CF}(\Theta_{CF})$ jointly with $l_{Demand}(\Theta_{Demand}; \Theta_{CF})$. Thus, $CF_{jt}$ in the benchmark model has different value from that in the proposed model.
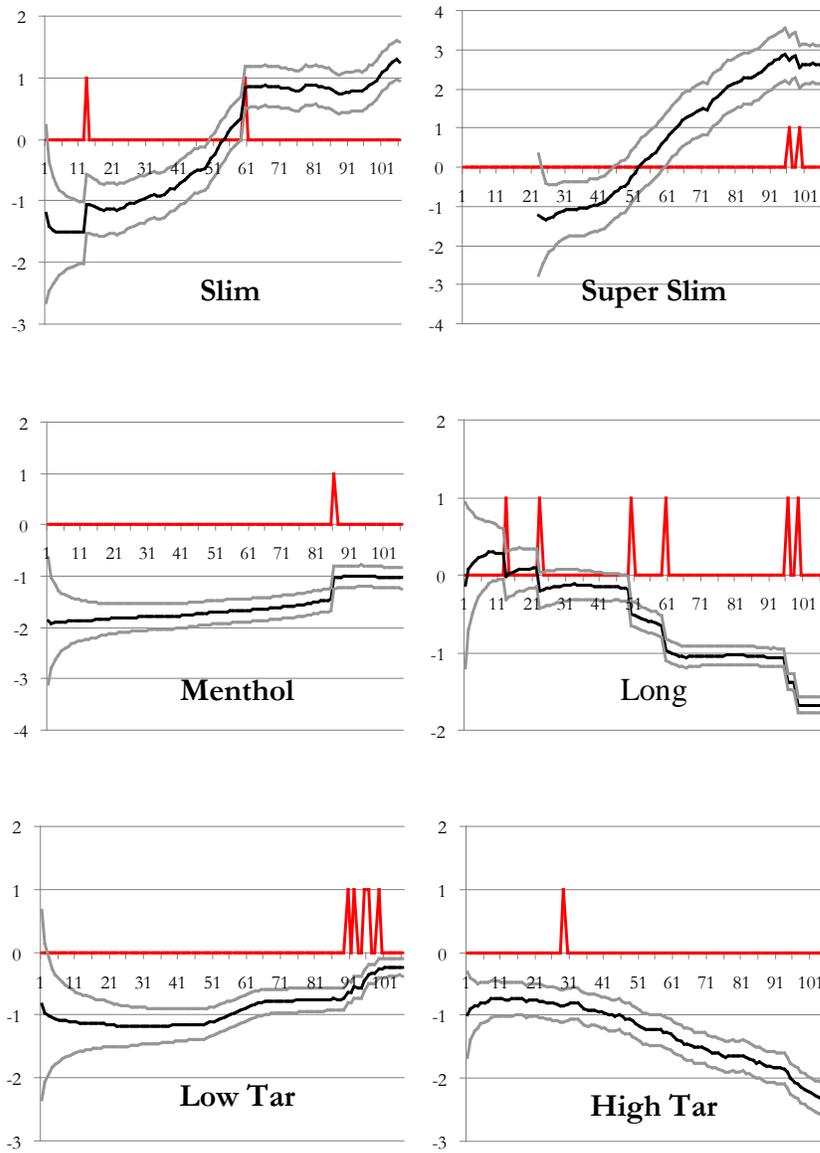
**Table 2.3: Estimation Results of Demand Model**

| | Proposed Model | | | | Benchmark Model | | |
|---|---|---|---|---|---|---|---|
| Parameter | | | | Parameter | | | |
| | Estimate | SE | t-value | | Estimate | SE | t-value |
| $\bar{\xi}$ | -2.686 | 0.198 | -13.6 | $\bar{\xi}$ | -2.072 | 0.141 | -14.7 |
| $\alpha$ (price) | -0.821 | 0.134 | -6.1 | $\alpha$ (price) | -1.222 | 0.087 | -14.1 |
| $\vartheta$ (CF) | -0.105 | 0.074 | -1.4 | $\vartheta$ (CF) | 0.160 | 0.046 | 3.5 |
| $\sigma_\xi$ | 1.008 | 0.017 | 58.6 | $\sigma_\xi$ | 1.079 | 0.017 | 64.8 |
| $\varphi^g$'s Slim | 0.480 | 0.224 | 2.1 | $\beta$ Slim | 0.297 | 0.123 | 2.4 |
| Super-slim | -0.223 | 0.369 | -0.6 | Super-slim | 1.395 | 0.381 | 3.7 |
| Menthol | 0.413 | 0.270 | 1.5 | Menthol | -1.475 | 0.171 | -8.6 |
| Long | -0.299 | 0.041 | -7.3 | Long | -0.719 | 0.100 | -7.2 |
| Low-tar | 0.098 | 0.040 | 2.5 | Low-tar | -0.539 | 0.113 | -4.8 |
| High-tar | -0.028 | 0.151 | -0.2 | High-tar | -1.203 | 0.088 | -13.7 |
| $\Sigma_\varsigma$ Slim | 0.083 | 0.021 | 3.9 | | | | |
| Super-slim | 0.178 | 0.041 | 4.4 | | | | |
| Menthol | 0.025 | 0.024 | 1.0 | | | | |
| Long | 0.002 | 0.013 | 0.2 | | | | |
| Low-tar | 0.007 | 0.017 | 0.4 | | | | |
| High-tar | 0.067 | 0.016 | 4.3 | | | | |
| $\Sigma_b$ Slim | 0.125 | 0.127 | 1.0 | $\Sigma_b$ Slim | 0.225 | 0.259 | 0.9 |
| Super-slim | 0.255 | 0.470 | 0.5 | Super-slim | 0.337 | 0.257 | 1.3 |
| Menthol | 0.212 | 0.274 | 0.8 | Menthol | 0.308 | 0.284 | 1.1 |
| Long | 0.087 | 0.073 | 1.2 | Long | 0.106 | 0.092 | 1.2 |
| Low-tar | 0.124 | 0.313 | 0.4 | Low-tar | 0.280 | 0.282 | 1.0 |
| High-tar | 0.285 | 0.166 | 1.7 | High-tar | 0.214 | 0.102 | 2.1 |
| $\sigma_\alpha$ | 0.049 | 0.028 | 1.7 | $\sigma_\alpha$ | 0.041 | 0.030 | 1.4 |
| P1000 | 0.349 | 0.079 | 4.4 | P1000 | 0.279 | 0.070 | 4.0 |
| P1500 | 0.133 | 0.108 | 1.2 | P1500 | 0.041 | 0.101 | 0.4 |
| P2000 | 0.058 | 0.151 | 0.4 | P2000 | 0.487 | 0.146 | 3.3 |
| PIA1 | -0.423 | 0.129 | -3.3 | PIA1 | -0.339 | 0.134 | -2.5 |
| PIA2 | -0.134 | 0.128 | -1.1 | PIA2 | -0.096 | 0.131 | -0.7 |
| PIB1 | 0.465 | 0.123 | 3.8 | PIB1 | 0.505 | 0.130 | 3.9 |
| PIB2 | 0.085 | 0.123 | 0.7 | PIB2 | 0.099 | 0.127 | 0.8 |
| Jan | -0.155 | 0.092 | -1.7 | Jan | -0.193 | 0.093 | -2.1 |
| Feb | -0.253 | 0.090 | -2.8 | Feb | -0.274 | 0.092 | -3.0 |
| Age | -0.120 | 0.186 | -0.6 | Age | -0.234 | 0.164 | -1.4 |
| Age^2 | 0.055 | 0.077 | 0.7 | Age^2 | -0.054 | 0.063 | -0.9 |
| 1/Age | -0.725 | 0.274 | -2.6 | 1/Age | -0.721 | 0.266 | -2.7 |
| LL | | -3083 | | LL | | -3164 | |
| AIC | | 6236 | | AIC | | 6385 | |
| BIC | | 6434 | | BIC | | 6549 | |

The estimates of $\varphi^g$'s and $\Sigma_\varsigma$ tell us about the evolution of consumer preferences for cigarette attributes. $\varphi^g$ captures the introduction effect. The introduction effects of "Slim" and "Low-tar" are significantly positive and the introduction effect of "Long" is significantly negative. The positive introduction effect of "Slim" means that after controlling for the influences of other factors (e.g. price and other demand shifters), the preference for slim products measured by the average sales of slim products compared to those of base products (i.e. regular thickness, 84 mm, non-menthol, medium-tar cigarette) increases after the introduction of a new slim product. The introduction effects of "Low-tar" and "Slim" can be interpreted similarly. The negative introduction effect of "Long" might be attributed to cannibalization. If the sales of a new slim product are due to a switch from existing slim products, the average sales of slim products compared to the average sales of base products decrease. This leads to the negative introduction effect.

The estimates of $\Sigma_\varsigma$ tell us that the exogenous effects of "Slim," "Super-slim," and "High-tar" are significant. Figure 2.4 shows the estimated $\beta_t^g$'s over time and 90% confidence bands. The changes in preferences are evident and statistically significant in all attributes except "Menthol." I observe increasing trends in preferences for "Slim," "Super-slim," and "Low-tar" and decreasing trends in preferences for "Long" and "High-tar". At the beginning of the sample period, slim and super-slim cigarettes are less preferred to regular-thickness cigarettes (note that preference for regular-thickness cigarette is normalized to zero). During the sample period, the preferences for these attributes have increases and at the end of the sample period slim and super-slim cigarettes are preferred to regular-thickness cigarettes. Long cigarettes are losing popularity during the sample period and become significantly less preferred to 84 mm regular-length cigarettes. Low-tar cigarettes become popular while high-tar cigarettes lose popularity. The estimation result implies that the increasing popularity of low-tar

cigarettes is attributed to the introduction effect while the decreasing popularity of high-tar cigarettes is attributed to the exogenous effect (e.g. increasing health consciousness). I also have significant estimates of the variances in heterogeneity distributions of prices and "High-tar" attribute.

The coefficient for $P_{1000}$ is significantly positive, supporting my hypothesis that consumers prefer round-priced products because of the transactional convenience. Consumers stockpile in the months before a price increase occurs and this is captured by significantly positive estimates of coefficient for *PIB1*. Also, the months in which price increases occurred show that sales decrease as the significantly negative coefficient for *PIA1* implies. Dummies for January and February are significantly negative and this indicates the significant decreases in demand during these months. This can be attributed to many smokers' unsuccessful attempts to quit smoking as the new year starts. The coefficient of $1/Age_{jt}$ is significantly negative and this implies that the demand of the newly introduced product is low during the first few months but increases subsequently and then plateaus.

**Figure 2.4: Time-Varying Preferences and 90% Confidence Bands**

x-axis: month; y-axis: preference; black line: $\beta_t^g$; grey line: 90% confidence band; red

line: $\eta_t^g$

### 2.5.2 Estimation Results of Pricing Model

Table 2.4 shows the estimation result of the proposed pricing model. It also reports the estimation result of a benchmark pricing model wherein we assume that the consumer preference is not varying over time or $\beta_t = \beta_{t-1}$ for all $t$. Note that under this assumption, (11) collapses to a linear regression of prices on time trend $t$ and product attributes $X_j$:

$$P_{jt} = \gamma_c + t \cdot \gamma_t + X'_j \gamma_x + \varepsilon^P_{jt}. \tag{25}$$

The model fit measures indicate that including the interaction of time-varying parameters and product attributes significantly improves the model fit after penalizing the increased model complexity. Except $\gamma_{mt}$, all the estimated coefficients for the interactions are significantly positive. This implies that when the preference for an attribute increases (decreases), the firm also increases (decreases) the prices of products with this attribute. This finding provides empirical evidence that KT&G strategically decides cigarette prices to benefit from changing consumer preferences. $\gamma_{lt}$ and $\gamma_{ht}$ are the greatest among the coefficients for the interactions, implying that the cigarette prices respond the most to changing preferences for tar contents.

As expected, $\gamma_t$ is significantly positive. This captures the price increases which have occurred during the sample period. The estimates of $\gamma_x$'s represent the average price level of products with each attribute level. The average price level of super-slim products is markedly higher than that of regular-thickness products ($\gamma_{x,ss} = 0.266$). Also, the average price level of low-tar products is higher than that of medium-tar or high-tar products ($\gamma_{x,lt} = 0.288$; $\gamma_{x,ht} = -0.057$). $\gamma_{x,mt}$ and $\gamma_{x,lg}$ are not significantly different from zero, implying that the average price levels of menthol products and long products are not different from those of non-menthol products and regular-length products.

**Table 2.4: Estimation Results of Pricing Model**

| Parameter | Proposed Model (time-varying preferences) | | | Benchmark Model (static preferences) | | |
|---|---|---|---|---|---|---|
| | Estimate | SE | t-value | Estimate | SE | t-value |
| $\gamma_c$ | 0.812 | 0.022 | 37.7 | 0.926 | 0.015 | 59.7 |
| $\gamma_t$ | 0.009 | 0.000 | 29.1 | 0.007 | 0.000 | 40.1 |
| $\gamma_{x,sl}$ | 0.098 | 0.019 | 5.2 | 0.137 | 0.019 | 7.2 |
| $\gamma_{x,ss}$ | 0.266 | 0.033 | 8.1 | 0.403 | 0.027 | 14.7 |
| $\gamma_{x,mt}$ | 0.051 | 0.078 | 0.7 | -0.070 | 0.017 | -4.0 |
| $\gamma_{x,lg}$ | 0.018 | 0.026 | 0.7 | -0.148 | 0.017 | -8.5 |
| $\gamma_{x,lt}$ | 0.288 | 0.041 | 7.1 | 0.107 | 0.019 | 5.7 |
| $\gamma_{x,ht}$ | -0.057 | 0.036 | -1.5 | -0.369 | 0.013 | -28.4 |
| $\gamma_{sl}$ | 0.063 | 0.018 | 3.6 | | | |
| $\gamma_{ss}$ | 0.093 | 0.017 | 5.5 | | | |
| $\gamma_{mt}$ | 0.079 | 0.050 | 1.6 | | | |
| $\gamma_{lg}$ | 0.174 | 0.025 | 7.0 | | | |
| $\gamma_{lt}$ | 0.288 | 0.050 | 5.8 | | | |
| $\gamma_{ht}$ | 0.223 | 0.024 | 9.3 | | | |
| R-square | | 0.65 | | | 0.62 | |
| Adjusted R-square | | 0.65 | | | 0.62 | |
| S.E. of regression | | 0.23 | | | 0.24 | |

### 2.5.3 Estimation Results of New Product Introduction and Attribute Choice Model

Along with the proposed model, I estimate a benchmark model wherein the firm believes that the consumer preferences are not changing over time and make its new product introduction and attribute selection decisions based on this belief. To be more specific, I plug in the static parameter demand model (the benchmark demand model with $\beta_t = \beta_{t-1}$) and pricing model (the benchmark pricing model (30)) to calculate the expected utilities and expected prices of the candidate products. From these, we calculate the expected profit and the introduction probability of each candidate product.

Table 2.6 reports the estimation results of the benchmark model and the proposed model. By comparing the log-likelihood of two models, we can conclude that the proposed model is markedly better than the benchmark model in explaining the firm's decisions regarding the new product introduction and attribute selection. Note that this comparison is equivalent to the formal likelihood-based test of firm behavior. It empirically supports the hypothesis that the firm knows that the consumer preferences are changing over time and it makes decisions based on the knowledge about varying preferences.

Figure 2.5 presents the predicted introduction probabilities of two candidate products: type-2 product ($k$=2) and type-15 product ($k$=15). Type-2 products are regular-thickness, regular-length, non-menthol, and medium-tar cigarettes. Note that 6 products out of 17 new products introduced during the sample period belong to this type. Type-15 products are super-slim, over 100 mm length, non-menthol, and low-tar cigarettes. The most recently introduced two products belong to this product type. In Figure 2.5, the numbers on the x-axis represent product types of actually introduced products during the sample period in time order. The predicted introduction

probabilities of the proposed model explain the variations in the introduced product types better than those of the benchmark model. When the type-2 product is introduced to the market, the predicted introduction probability of the type-2 product is high. When the type-15 product is introduced to the market, the predicted introduction probability of the type-2 product decreases and that of the type-15 product increases. In contrast, the predicted introduction probability of the type-2 product is always higher than that of the type-15 product in the benchmark model. The variations in the introduced product types are explained by varying consumer preferences in the proposed model, along with other factors (e.g. cannibalization effect, price changes, and cost factors). On the other hand, the benchmark model does not consider the time-varying preferences in the explanation of the variation and, as a result, shows poor performance in the explanation of the firm decision regarding the new product introduction and attribute choice. The key implication here is that the firm monitors consumers' time varying preferences for product attributes and consider these in its decisions regarding new product introduction and selection of attribute-levels for product design. The results of the demand model imply that the firm's action (i.e. introduction of a certain product) subsequently influences consumer preferences for product attributes via the introduction effect, along with the exogenous effects (e.g. social trends). In conclusion, the market evolves as an outcome of the interaction between the firm's marketing activities (i.e. new product design and introduction) and changes in consumer preferences.

**Table 2.5: Estimation Results of New Product Introduction and Attribute Selection Model**

| Parameter | Proposed Model: Myopic Firm (time-varying preferences) | | | Benchmark Model (static preferences) | | |
|---|---|---|---|---|---|---|
| | Estimate | SE | t-value | Estimate | SE | t-value |
| $\sigma_V$ | 0.096 | 0.034 | 2.8 | 0.176 | 0.094 | 1.9 |
| $c_{x,sl}$ | -0.097 | 0.210 | -0.5 | -0.177 | 0.330 | -0.5 |
| $c_{x,ss}$ | 0.262 | 0.157 | 1.7 | 0.090 | 0.286 | 0.3 |
| $c_{x,mt}$ | 0.132 | 0.464 | 0.3 | 0.471 | 0.724 | 0.7 |
| $c_{x,lg}$ | 0.222 | 0.181 | 1.2 | 0.320 | 0.323 | 1.0 |
| $c_{x,lt}$ | -0.104 | 0.061 | -1.7 | 0.092 | 0.152 | 0.6 |
| $c_{x,ht}$ | -0.028 | 0.190 | -0.1 | 0.088 | 0.281 | 0.3 |
| $F$ | $3.96 \times 10^6$ | $1.31 \times 10^6$ | 3.0 | $6.67 \times 10^6$ | $3.36 \times 10^6$ | 2.0 |
| LL | -76.3 | | | -82.6 | | |
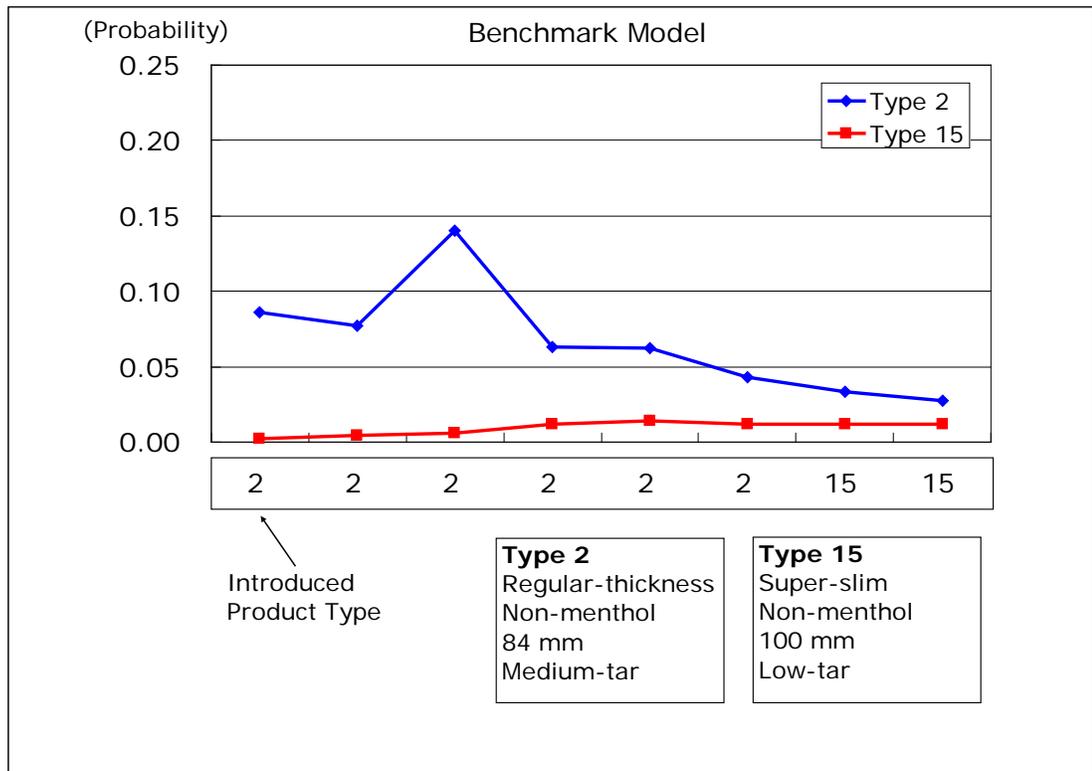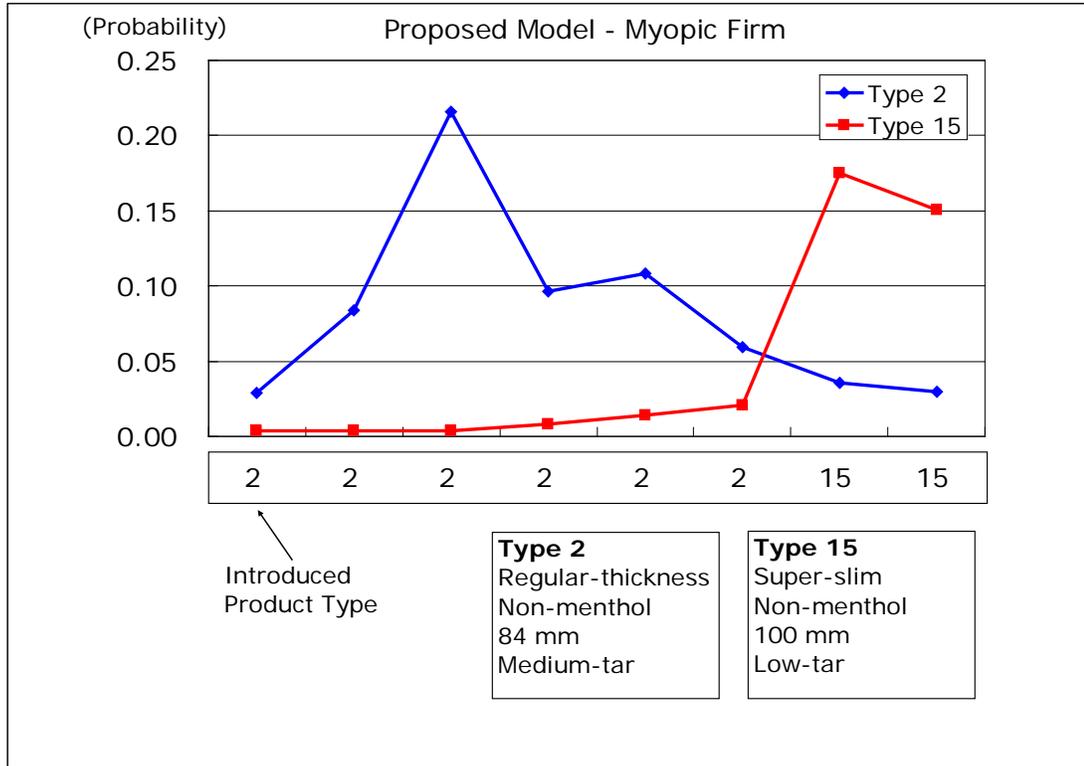
**Figure 2.5: Predicted Introduction Probabilities of Selected Candidate Products**

The estimates of $c_x$ provide us with the information on the marginal cost. The estimate of $c_{x,ss}$ is significantly positive, implying that the marginal costs of the super-slim cigarettes are higher than slim or regular-thickness cigarettes. The estimate of $c_{x,lt}$ is significantly negative, implying that the marginal costs of the low-tar cigarettes are lower than other products. To make low-tar cigarettes, manufacturers usually mix the expanded tobacco leaf or tobacco stalk with regular tobacco leaves. Considering that the expanded tobacco leaf and tobacco stalk costs much less than regular leaves, my result of low marginal cost for low-tar product can be explained. The estimated $F$ tells us the fixed cost of adding one product to the existing product line. The estimate amounts to 3.66 billion KRW.

## 2.6 Conclusion

In this study, I proposed a model of market evolution. On the demand side, the market evolution is characterized by changes in consumers' preferences for product attributes. On the supply side, the market evolution is characterized by the firm's new product design and introduction. The demand and the supply sides interact with each other over time and the market evolves as a result. The proposed structural model describes the details of market evolution by capturing the interaction between the demand and the supply sides. Consequently, it provides a deeper understanding of the market evolution. To summarize the substantive findings from the empirical analysis of the South Korean cigarette market using the proposed model,

i. During the 106-month sample period from January 1995 to October 2003, consumers' preferences for cigarette attributes show significant change over time. To be more specific, preference for the slim attribute increases due to both the introduction effect and the exogenous effect. Preference for the super-slim attribute increases due to the exogenous effect. Preference for the long attribute decreases due to the introduction effect. Preference for the low-

92

tar attribute increases due to the introduction effect. Preference for the high-tar attribute decreases due to the exogenous effect. Preference for the menthol attribute does not show significant change. Significant introduction effects imply that the firm can influence consumers' preference by introducing and designing new products. This finding sheds light on the role and value of new product design and introduction in a market where consumer preferences are evolving, and the firm's ability to influence demand through prices or advertising is constrained by regulation.

ii. Preference change explains the substantial variation in price change over time. The time-varying preference for an attribute is positively correlated with the prices of cigarettes which have the attribute. This implies that when the preference for an attribute increases (decreases), the firm also increases (decreases) the prices of products with this attribute. This finding provides empirical evidence that KT&G strategically decides cigarette prices to benefit from changing consumer preferences.

iii. If a firm strategically sets cigarette prices based on its knowledge of time-varying preferences, a static demand model may suffer from endogeneity due to the correlation between price and the error term which contains the unexplained variance from the time-varying preferences. My empirical analysis illustrates the seriousness of this endogeneity problem. The estimated price coefficient is -0.821 in the proposed demand model wherein consumers' preferences for cigarette attributes are varying over time. When I do not allow consumer preferences to change, the price coefficient becomes -1.222, a significant increase of 49%. The price elasticity of cigarette demand has been extensively studied in the cigarette-related literature and has played a prominent role in legislative debates about using taxation as a principal tool to

discourage smoking. My result shows that proper consideration of this is necessary for precise estimation of the price elasticity if consumer preferences are changing over time.

iv. The proposed new product introduction and attribute choice model explains the observed data much better than the benchmark model. This empirically supports the hypothesis that the firm knows that the consumer preferences are changing over time and it makes decisions based on the knowledge about varying preferences. The variations in the introduced product types are explained by varying consumer preferences in the proposed model whereas the benchmark model does not consider the time-varying preferences in the explanation of the variation. We find that the marginal costs of the super-slim cigarettes are higher than slim or regular-thickness cigarettes and the marginal costs of the low-tar cigarettes are lower than other products. Also, the estimated fixed cost of adding one product to the existing product line is 3.66 billion KRW.

Methodologically, this study identifies a new slope endogeneity problem due to permanent shocks or time-varying consumer preferences. Also, it extends the aggregate random coefficient logit model to incorporate stochastically varying parameters in a likelihood-based framework.

Several directions exist for further research. In this study, we model the firm as a myopic agent. To be more specific, the firm considers its current profit for the decisions regarding its new products. However, the firm might consider its future profits as well in its decision makings. Also, the firm might take account of the impact of introduction on consumers' future preferences because the introductions shape consumer preferences through the introduction effects. Consequently, the forward-looking firm may solve a dynamic optimization problem. To incorporate the

forward-looking behavior of the firm into the current framework, one should modify the new product introduction and attribute choice model (Note that the demand model and the pricing model are consistent with the forward-looking firm). One challenge here is the "curse of dimensionality". The problem has a huge state-space and computational handling of this problem is challenging.

The South Korean cigarette market had been a de facto monopoly market during the sample period. Thus, in this study, I build my model assuming that the KT&G is a monopolist. To apply the proposed modeling framework to other industries, proper modifications should be made according to the structure of the industry. In particular, the pricing model and the new product introduction and attribute choice model should be modified appropriately. Moreover, if the firms are forward-looking, the supply side model becomes a form of dynamic game. Even though estimation of such a model will be challenging, it will provide rich implications to both practitioners and researchers.

In this study, I focus on the firm's decisions regarding its new products in a market where the consumers' preferences are changing over time. To the best of my knowledge, little attention is paid to the firm's decision regarding advertising and promotion when consumers' preferences are varying over time. Again, such research will provide rich implications to both practitioners and researchers.

REFERENCES

Ackerberg, D., C. L. Benkard, S. Berry, A. Pakes. (2007) Econometric Tools for
    Analyzing Market Outcomes. in Handbook of Econometrics, ed. J. J. Heckman
    and E. E. Leamer, vol. 6, part 1, Amsterdam, North Holland.

Becker, Gary S., Kevin M. Murphy (1988) A Theory of Rational Addiction, Journal of
    Political Economy, Vol. 96, No. 4, pp. 675-700.

Berry, S. (1994) Estimating Discrete-choice Models of Product Differentiation. RAND
    Journal of Economics. 25 242–262.

Berry, S., J. Levinsohn, A. Pakes. (1995) Automobile Prices in Market Equilibrium,
    Econometrica, Vol. 63, pp. 841–890.

Chaloupka, Frank. J., Kenneth E. Warner (2000) The Economics of Smoking in
    Handbook of Health Economics ed. A.J. Culyer and J.P. Newhouse, Elsevier
    Science.

Chen, Tao., Baohong Sun, Vishal Singh (2007) An Empirical Investigation of the
    Dynamic Effect of Marlboro's Permanent Pricing Shift, Marketing Science,
    forthcoming

Chong, Juin-Kuan., Teck-Hua Ho., Christopher S. Tang (2001) A Modeling
    Framework for Category Assortment Planning, M&SOM, Vol. 3, No. 3, pp.
    191-210

Dobson, Gregory., Shlomo Kalish. (1988) Positioning and Pricing a Product Line,
    Marketing Science, Vol. 7, No. 2, pp. 107-125

Draganska, Michaela., Michael Mazzeo, Katja Seim (2009) Beyond Plain Vanilla:
    Modeling Joint Product Assortment and Pricing Decisions, Quantitative
    Marketing and Economics, Vol. 7, No. 2, pp. 105-146

Dubé, Jean-Pierre., K. Sudhir., Andrew Ching., Gregory S. Crawford., Michaela

Draganska., Jeremy T. Fox., Wesley Hartmann., Günter Hitsch., V. Brian Viard., Miguel Villas-Boas., Naufel Vilcassim. (2005) Recent Advances in Structural Econometric Modeling: Dynamics, Product Positioning and Entry, Marketing Letters, Vol. 16(3), pp. 209–224.

Feldman, Eric A. (2001) The Landscape of Japanese Tobacco Policy: Law, Smoking and Social Change, Eric A. Feldman, The American Journal of Comparative Law, Vol. 49, No. 4, pp. 679-706

Green, Paul E., Abba M. Krieger. (1985) Models and Heuristics For Product Line Extension, Marketing Science, Vol. 4, No. 1, pp. 1-19

Hamilton, James D., (1989) A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle, Econometrica, Vol. 57, No. 2, pp. 357-384

Hitsch, Günter. (2006) An Empirical Model of Optimal Dynamic Product Launch and Product Exit Under Demand Uncertainty, Vol. 25(1), pp. 25-50.

Holak, Susan L., Srinivas K. Reddy (1986) Effects of a Television and Radio Advertising Ban: A Study of the Cigarette Industry, Journal of Marketing, Vol. 50, pp. 219-227

Horsky, Dan., Paul Nelson. (1992) New Brand Positioning and Pricing in an Oligopolistic Market, Marketing Science, Vol. 11(2), pp. 133–153.

Jiang, Renna, Puneet Manchanda, Peter E. Rossi (2009) Bayesian Analysis of Random Coefficient Logit Models Using Aggregate Data, Journal of Econometrics, (in press)

Kadiyali, Vrinda., Naufel Vilcassim., Pradeep Chintagunta. (1999) Product Line Extensions and Competitive Market Interactions: An Empirical Analysis, Journal of Econometrics, Vol. 89, pp. 339-363

Keane, M. P. 1993. Simulation Estimation for Panel Data Models with Limited

Dependent Variables. C. R. Rao, G. S. Maddala, D. Vinod, eds. The Handbook of Statistics & Econometrics. Elsevier Science Publishers, Amsterdam, The Netherlands, 545–572.

Kim, C.-J. (2004) Markov-switching Models with Endogenous Explanatory Variables, Journal of Econometrics, 122(1), 127-136.

Kim, C.-J. (2006), Time-varying Parameter Models with Endogenous Regressors, Economics Letters, 92(1), 21-26.

Kim, Songjune., Barry J. Seldon (2004) The Demand for Cigarettes in the Republic of Korea and Implications for Government Policy to Lower Cigarette Consumption. Contemporary Economic Policy, 22(2), 299-308.

Kuksov, Dmitri and J. Miguel Villas-Boas (2008) Endogeneity and Individual Consumer Choice, Journal of Marketing Research, 45(6), pp. 702-714.

Luan, Y. Jackie., K. Sudhir (2009) Forecasting Marketing Mix Responsiveness for New Products, Journal of Marketing Research, forthcoming

Machado, Fernando S., Rajiv K. Sinha. (2007) Smoking Cessation: A Model of Planned vs. Actual Behavior for Time-Inconsistent Consumers, Marketing Science, Vol. 26(6), pp. 834-850.

Mazzeo, Michael. (2002) Product Choice and Oligopoly Market Structure, RAND Journal of Economics, Vol. 33, pp. 221–242.

Min, Heechul. (2007) Cigarette Demand Estimation Using Discrete Choice Model, Monthly Public Finance Forum (in Korean) (138), p. 27-42

Musalem, Andrés., Eric T. Bradlow, Jagmohan S. Raju (2009) Bayesian Estimation of Random-Coefficients Choice Models Using Aggregate Data, Journal of Applied Econometrics, Vol. 24, pp. 490–516

Nevo, Aviv. (2001) Measuring Market Power in the Ready-to-Eat Cereal Industry, Econometrica, Vol. 69(2), pp. 307-342.

Park, Sungho., Sachin Gupta (2008) Comparison of SML and GMM Estimators for the Random Coefficient Logit Model Using Aggregate Data, working paper, Cornell University

Park, Sungho., Sachin Gupta (2009) A Simulated Maximum Likelihood Estimator for the Random Coefficient Logit Model Using Aggregate Data, Journal of Marketing Research, forthcoming

Petrin, Amil., Kenneth Train (2009) A Control Function Approach to Endogeneity in Consumer Choice Models, Journal of Marketing Research, forthcoming

Reddy, Srinivas K., Susan L. Holak., Subodh Bhat (1994) To Extend or Not to Extend: Success Determinants of Line Extensions, Journal of Marketing Research, Vol. 31 (May), pp. 243-262

Shen, Qiaowei (2008) Industry Evolution with Endogenous Entry and Exit, Working Paper, The Wharton School, University of Pennsylvania.

Shin, Yoon-Jung (2007) Analysis of deceptiveness and health hazard for tobacco contents and product types, Report 2007-28, Korea Institute for Health and Social Affairs.

Shocker, Allan D., V. Srinivasan (1979) Multiattribute Approaches for Product Concept Evaluation and Generation: A Critical Review, Journal of Marketing Research, Vol. 16 (May), pp. 159-180

Sriram, S., Pradeep Chintagunta, and Ramya Neelamegham (2006) Effects of Brand Preference, Product Attributes, and Marketing Mix Variables in Technology Product Markets, Marketing Science, Vol. 25, pp. 440-456

Sudhir, K. (2001) Competitive Pricing Behavior in the US Auto Market: A Structural Analysis, Marketing Science, Vol. 20(1), pp. 42-60.

Villas-Boas, J. Miguel., Russell S. Winer. (1995) Endogeneity in Brand Choice Models, Working paper, University of California at Berkeley

Villas-Boas, J. Miguel., Russell S. Winer. (1999) Endogeneity in Brand Choice

Models, Management Science, Vol. 45(10), pp. 1324-1338.

# CHAPTER 3

# A REGIME-SWITCHING MODEL OF CYCLICAL CATEGORY BUYING

## 3.1 Introduction

> "*I think I am seriously addicted to Greek yogurt. For a while,*
>
> *I've been having it every day for breakfast. It's just so good.*
>
> *…… Anyway, the point is that I'm a little worried that having*
>
> *the same thing for breakfast every day is not good for me. So I*
>
> *tried having orange-carrot juice and an apple for breakfast.*
>
> *I'm going to try the juice thing or something else for a while*
>
> *though, and maybe later I'll come back to Greek yogurt. It*
>
> *could happen.*"

The breakfast diary of an anonymous consumer indicates a continuous period of consumption of Greek yogurt, then orange-carrot juice, and then Greek yogurt again. This pattern of behavior is not atypical for frequently purchased, nondurable goods -- many readers will recognize similar patterns in their own everyday life. For some period of time, consumers are strongly inclined to use or consume a certain product repetitively. This "obsession" with the product then ends due to various reasons (discussed subsequently), and possibly resumes again after a hiatus.
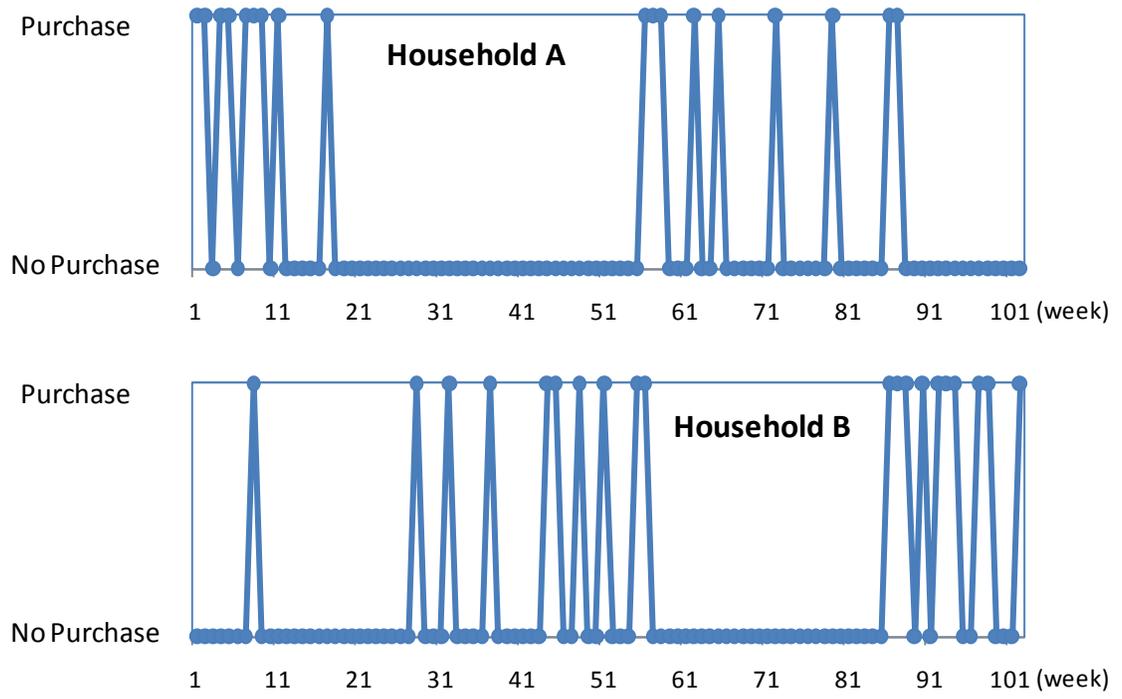
**Figure 3.1: Examples of Cyclical Yogurt Purchasing Behavior in IRI Households**

The consumption behavior of the anonymous consumer quoted above resembles the purchase history of yogurt and other products in the recently released IRI household panel data (Bronnenberg et al. 2008)[33]. In Figure 3.1 we show the weekly yogurt purchase behavior of two panelist household during a two-year period. "Purchase" and "No Purchase" represent weeks with and without a yogurt purchase, respectively. In the upper panel, panelist household 'A' repeatedly purchases yogurt for the first 17 weeks and then does not purchase yogurt for 38 weeks. Recognizing that yogurt is very perishable, we may reasonably infer from the long period of not purchasing that the household did not consume any yogurt for most of this period. This no-purchase period is followed by another period of frequent yogurt purchasing, which is followed by another period of not purchasing. In the lower panel, panelist household 'B' shows a similar cyclical purchasing pattern.

How prevalent are such cyclical purchasing patterns? Put otherwise, what proportion of yogurt purchasing households display long periods of non-purchase and non-consumption? To answer this question we look in greater detail at household behavior in the IRI data. Since in these data we can only observe purchasing, and not consumption behavior, we consider three product categories that are more perishable – yogurt, milk, and frankfurters – so that we may be able to infer periods of non-consumption based on periods of non-purchasing.

We first define the terms "Purchase Week", "No Purchase Week", and "No Consumption Period" for the yogurt category (analogous definitions apply to the other two categories). If a household purchases any product in the yogurt category in a certain week, that week is labeled a "Purchase Week" and if not, a "No Purchase

---

[33] While we describe these data in greater detail in Section 3.4, where we use the same data for our empirical analysis with the proposed model, it is worth noting at this point that the data capture panelist households' shopping trips and yogurt purchasing in *all* stores in all retail channels, including grocery, drug, mass merchandise, club, convenience stores and specialty stores.

Week." "No Consumption Period" (NCP) is defined as a period of consecutive "No Purchase Weeks" that lasts ten or more weeks. Yogurt products usually show a "best by" date on the container. We conducted a field survey of this date which revealed that the average "best by" (or "sell by") date is 28 days or four weeks (standard deviation = 8 days) from the date of purchase (the day we performed the field survey).[34] Although households may sometimes consume yogurt after the "best by" or "sell by" date, we believe it is reasonable to conjecture that <u>ten</u> or more weeks of no yogurt purchase implies a substantial period of *no consumption*.

In IRI's household panels in Eau Claire, Wisconsin and Pittsfield, Massachusetts, over a period of 104 weeks in 2003 and 2004, 4,298 panelist households (HHs) purchased yogurt (excluding drinking yogurt) at least once (i.e., their count of "Purchase Week" is at least one). Table 3.1 shows the distribution of NCPs in the data. Of the 4,298 HHs, 73% show one or more NCPs, and 49% show two or more NCPs. The average length of NCPs in the data is 19.7 weeks. To get insights into households that buy yogurt more frequently, we consider the 1,559 HHs whose count of "Purchase Week" is 20 or more[35]. Of these HHs, 62% show one or more NCPs, and 33% show two or more NCPs.

Similar statistics are provided for the milk and frankfurters categories in Table 3.1. Based on the "best by" dates that we observed in these categories in our field survey, the "No Consumption Period" (NCP) is defined as a period of consecutive "No Purchase Weeks" which lasts six or more weeks for milk, and eighteen or more weeks for frankfurters[36].

---

[34] We visited three grocery stores belonging to three different chains on the East Coast.
[35] The average count of "Purchase Weeks" in 4,298 HHs who purchased yogurt during the sample period is about 20.
[36] Our survey revealed that the average "best by" date is 19 days from date of purchase in milk, and 50 days in frankfurters.

The simple analysis presented in Table 3.1 reveals an important facet of consumer purchase and consumption behavior in the three categories. A substantial proportion of households show cyclical category purchase behavior, which in these categories also implies cyclical category consumption behavior. In each of the three categories, over two-thirds of buying households display at least one episode of several consecutive weeks of no-purchase. This episode is long enough that we can be reasonably certain there is non-consumption as well. Moreover, this phenomenon is not limited to infrequent category buyers. Over half of frequent buyers in the three categories also display at least one such episode.

A possible alternative explanation of the data in Table 3.1 is that households were temporarily outside the market area due to, for instance, being on vacation, hence we do not see yogurt being bought. Our analysis offers considerable evidence against such a vacation theory. First, if a household was on vacation outside the market area, we should not observe any visits to stores in the area. Defining a "store visit week" as a week in which a household visits any area store, regardless of yogurt purchase, we show in Table 3.2 the distribution of this variable *during* No Consumption Periods for each of the three categories. The data indicate that on average, a store visit occurred in about 80% of the weeks in a NCP. Thus, it is unlikely that households were away from the market area for most of the duration of a NCP.

**Table 3.1: Distribution of "No Consumption Period" (NCP) in Three Categories**

| | All Category Buyers | | | Frequent Category Buyers* | | |
|---|---|---|---|---|---|---|
| | Yogurt | Milk | Frankfurter | Yogurt | Milk | Frankfurter |
| % of HHs with #NCPs ≥1 | 73% | 71% | 66% | 62% | 56% | 58% |
| % of HHs with #NCPs ≥2 | 49% | 54% | 21% | 33% | 32% | 14% |

* Frequent Category Buyers are defined as HHs whose count of Purchase Weeks is at least 20 in yogurt, 40 in milk, and 9 in frankfurters, over a 104-week period. These thresholds are the mean counts of Purchase Weeks in the respective categories.

Second, for yogurt, the distribution of number of NCPs ending in each calendar week shows a relatively constant pattern over the time period (with a slight peak in January).  Since we might expect certain periods of the year to be more popular for vacation, this evidence is also inconsistent with the vacation theory.  Third, the average length of NCPs in yogurt is 19.7 weeks, which is substantially longer than most vacations.  Finally, as we describe in more detail subsequently, our model is based on store visits being classified as belonging to high versus low purchase tendency.  If in certain periods *no store visits occur* because the household is away from the market area, these periods are effectively excluded in the analysis.  This implies that our model-based findings cannot be due to the presence of vacation periods in the data.

These episodes of cyclical consumption may be explained by a cognitive need for stimulation in the context of exploratory and novelty-seeking behavior (Berlyne 1970; Roger 1979).  In particular, Berlyne's theory of exploratory behavior proposes that the attractiveness of a stimulus is an inverted U-shaped function of its familiarity.  This relationship is explained as the joint effect of habituation and boredom.  Habituation comes into play when consumers are exposed to a relatively unfamiliar stimulus or product.  Due to the habituation factor, consumers tend to repeat the stimulus and thereby increase familiarity with it until its attractiveness is maximized.  On the other hand, high familiarity also arouses boredom and the total attractiveness decreases as the stimulus repeats.  According to this theory, high and low purchase tendencies can be attributed to the habituation and boredom factors, respectively.

**Table 3.2: Distribution of "Store Visit Weeks" During No Consumption Periods**

|  | *Yogurt* | *Milk* | *Frankfurters* |
|---|---|---|---|
| Number of NCPs | 6,793 | 11,172 | 3,979 |
| Average length of NCP | 19.7 weeks | 10.2 weeks | 30.6 weeks |
| Average count of "Visit Weeks" *during* NCPs | 16.8 weeks | 7.9 weeks | 26.4 weeks |

There are several mechanisms that may lead consumers to keep consuming and purchasing a product, such as satisfaction obtained from consuming the product, or from the consumer's need to routinize behavior so as to minimize the cost of thinking (Jeuland 1979; Shugan 1980). Considerable research in psychology recognizes that much of everyday action is characterized by habitual repetition (Neal, Wood, and Quinn 2006). If consumers repeatedly consume a product, they may become satiated, at least temporarily, and prefer to consume another product (Coombs and Avrunin 1977; McAlister 1982). This explains the ending of a period of persistent consumption of a category and the beginning of the period of low purchase tendency.

In addition to internal factors (e.g. satisfaction, boredom, satiation), external factors (e.g. news, introduction of a new product, price changes, advertising or promotions, and seasonal availability) may also explain consumers' switching behavior between high and low purchase tendency episodes. For example, in June 2008 the US Food and Drug Administration (FDA) recommended that consumers should avoid eating raw tomatoes due to fears of salmonella. A consumer who had a high inclination to consume raw tomatoes at this time is likely to cease purchasing and consuming raw tomatoes in response to this news. Instead, tomato juice or tomato soup might be consumed persistently as the alternative tomato-delivery vehicle of choice. A month later, the FDA declared that tomatoes were safe to eat. With this update, the consumer may revert to buying raw tomatoes.

Our research goal in the present paper is to propose a descriptive model of the kind of cyclical purchasing behavior for which we have provided evidence. As discussed, alternative theoretical explanations might be consistent with such behavior. Our goal is not to disentangle these explanations. Testing the alternative theories empirically would require richer data than we have access to, as well as the use of structural econometric models that impose greater structure on consumer behavior.

We leave these developments to future research. Here we propose a Markov regime-switching random coefficient logit model to represent switching behaviors between high and low purchase tendencies in consumers' purchase decisions in a category. Consumers are assumed to be heterogeneous in their intensities of high and low purchase tendencies. In addition to category purchase incidence, we also consider consumers' brand choice decisions and the impact of marketing mix variables within the random coefficient logit framework. Since the proposed model nests the typical random coefficient logit model without Markov regime-switching parameters, we can test if purchase behavior can be better characterized by allowing for switching category purchase tendencies.

The main feature of the proposed model is that it divides the stream of purchase decisions of each consumer into distinct regimes with different parameter values that characterize high and low purchase tendencies. Specifically, we introduce a regime-switching intercept in the consumer's indirect utility function. We interpret the regime-switching intercept as a proxy for the consumer's inventory in categories which are substitutes for the target category, i.e., the category of interest. For instance, the consumer in our opening quote tends to consume orange-carrot juice or apple (substitute categories) during periods when the consumer has low purchase tendency for yogurt (the target category). The regime-switching intercept approximates the level of substitute inventory, which is unobservable to the researcher but can be inferred from the consumer's purchasing behavior.

This interpretation of the regime-switching intercept highlights an important methodological concern with models of category purchase incidence. We expect that the target category inventory and substitute category inventory are negatively correlated, and both influence the consumer's purchase decision in the target category. As a consequence, the omission of substitute category inventory may cause an

endogeneity problem. This problem is widespread because, to our knowledge, almost all applications of random utility models to category buying ignore substitute category inventory. We investigate the problem of endogeneity of target category inventory empirically and demonstrate how the proposed model handles it.

In addition to including heterogeneity across consumers and parameter dynamics (regime switching), the proposed model also incorporates unmeasured product characteristics (i.e. common shocks) and considers the endogeneity of prices. The importance of these issues has been extensively documented in the literature (Berry 1994; Villas-Boas and Winer 1999; Chintagunta et al. 2005). To handle the endogeneity of prices, we use the control function method (Petrin and Train 2004; Park and Gupta 2008). The resulting Markov regime-switching random coefficient logit model accounts for parameter dynamics, heterogeneity across consumers, endogeneity of prices, and endogeneity of inventory. In sum, the proposed model incorporates several important aspects of consumer choice in as complete a manner as possible.

We apply the proposed model to yogurt purchases of a sample of yogurt-buying households and find that as many as 40.8% of the households display cyclicality in buying, after controlling for the effects of marketing mix variables, state-dependence in brand choice, and inventory. Predictions from the proposed model track observed purchases of households closely, and the model also fits better than three benchmark models. We show that if the model ignores the underlying dynamics of switching between high and low purchase tendencies, an endogeneity problem arises.

We also show that cyclicality in buying has a key implication for a firm's price promotion strategies: a price reduction that is offered to a household during its high purchasing tendency period results in greater increases in sales than one that is

offered during its low purchasing period. This opens up the opportunity for customized timing of price reductions as a new dimension for enhancing the effectiveness of promotions. We show via simulation that in our sample data, a one-time 30% price reduction on Yoplait with customized timing leads to a 92% increase in the impact of the promotion, relative to a randomly timed promotion. Finally, since the proposed model can be challenging for firms to estimate, we explore the use of simple descriptive statistics of consumers' purchasing behavior (mean and standard deviation of inter-purchase times) to classify households into groups based on their cyclicality in category buying. Further, a simple rule-of-thumb also based on past purchasing allows the firm to guess quite successfully whether a household is in the high or low purchase state. These findings reveal an opportunity firms to improve the efficiency of their promotions without estimating the proposed model.

The remainder of the paper is organized as follows. In the next section, we review the existing relevant literature. In section 3.3, we present the model and explain our estimation method. In section 3.4, we describe an empirical application of the model to scanner panel data and discuss our key findings and their managerial implications. Finally, in section 3.5, we summarize the contributions of this article and identify future research issues.

**3.2 Literature Review**

We summarize two sets of literature relevant to our research. The first set includes work on serial dependence in category purchase incidence and brand choice decisions. In the second set we include a brief review of work on heterogeneity, endogeneity, and time varying parameters in choice models.

**3.2.1 Serial dependence in category purchase incidence and brand choice decisions**

In this paper, we investigate switching behaviors between high and low purchase tendencies in consumers' category purchase decisions. We define high and low purchase tendencies in terms of conditional purchase probabilities (conditional on marketing mix variables, inventory, and unmeasured product characteristics). Since it is not possible to observe whether a consumer has a high purchase tendency or a low purchase tendency at time $t$, we infer this from the consumer's purchase decisions using latent variables. Furthermore, we expect high or low purchase tendencies to persist for some period of time, as illustrated by the aforementioned quote from the consumer's diary. In other words, time period $t$ tends to be characterized as having a high purchase tendency with higher probability than a low purchase tendency, if period $t$-$1$ is characterized as one with high purchase tendency. A similar serial dependence is expected for low purchase tendencies. We capture this serial dependence using a first-order Markov process.

In the literature, serial dependence of *category purchase decisions* has attracted little attention. A rare example is Ailawadi and Neslin (1998) who use an incidence, choice, and quantity model framework to investigate the effect of promotions on household category consumption. In the purchase incidence model specification, they include a lagged purchase incidence indicator variable to capture systematic swings in purchase and consumption due to "eating bouts, binging, special diets, and other situational factors." They expect that category purchase on one shopping trip might be associated with a higher likelihood of purchase on the next trip as a result of these phenomena. Results of an empirical application to the yogurt and ketchup categories show a significant positive estimate for the coefficient of the lagged incidence variable, consistent with their expectations. In section 3.3, we compare this approach (i.e. including a lagged incidence variable) with the proposed model and also use it as a benchmark in the empirical application in section 3.4.

In the context of brand choice, marketing researchers have long been interested in the effects of lagged brand choice (i.e. structural state dependence) and lagged brand choice probability (i.e. lagged evaluation effect or habit persistence) (Seetharaman 2003; Keane 1997; Roy et al. 1996; Heckman 1981). This *serial dependence of brand choice decisions* is often referred to as *state dependence*. In the random utility framework, researchers typically accommodate state dependence by introducing a lagged purchase variable (Jones and Landwehr 1988; Krishnamurthi and Raj 1991; Seetharaman et al. 1999), a variable constructed from lagged purchases (Erdem 1996; Guadagni and Little 1983; Keane 1997), or serial correlation in the error term (Keane 1997; Seetharaman 2003). The general finding in the literature is that low-priced, frequently purchased grocery categories are characterized by inertia (Erdem 1996; Keane 1997; Roy et al. 1996; Seetharaman and Chintagunta 1998), and that households display similar state dependence effects in their purchasing behavior across multiple categories (Seetharaman et al. 1999).

An interesting question arises here: can the proposed model be applied to the context of brand choice decisions to capture periods of high purchase tendency for a certain brand. After repeatedly purchasing a brand, a consumer may become satiated and prefer to try another brand. Note that we can explain this behavior with the same mechanisms that we offered to explain switching behaviors in category purchasing. Bawa (1990) explored the possibility that consumers may fluctuate between inertial and variety-seeking behavior at different times depending on their choice history. He accommodated such hybrid behavior in a brand choice model by specifying utility derived from a brand as a quadratic function of the number of consecutive purchases of that brand. His empirical analysis of purchasing in three product categories – facial tissue, paper towels, and ready-to-eat cereal – showed that more than half the sample households switched between inertia and variety seeking. Similarly, Ratner et al.

(1999) found that consumers alternate between habit-persistent and variety seeking states. In section 3.3, we specify a similarly motivated model (Benchmark 3) and compare it with the proposed model in section 3.4.

**2.2 Heterogeneity, endogeneity, and time-varying parameters in choice model**

Accounting for heterogeneity across households has become a crucial factor in estimating choice models. Accommodating heterogeneity enables the model to represent more realistic consumer choice behavior. Furthermore, it is now well established that failure to control for heterogeneity can result in biases in the estimated mean parameters of marketing-mix variables (Chintagunta et al. 1991; Allenby and Rossi 1999).

Along with heterogeneity, the marketing literature has recognized another source of bias in the estimation of choice models. If there are unmeasured product characteristics that influence consumer choices but are not observed by the researcher, while the marketer observes these and incorporates them into decision making, then an endogeneity problem could arise (Berry 1994; Villas-Boas and Winer 1999). Numerous empirical applications have shown that unmeasured product characteristics and price can be highly correlated in practice (Petrin and Train 2004).

Besides the endogeneity issue, unmeasured product characteristics, if omitted from the model, cause biases. Chintagunta et al. (2005) noted an upward bias in heterogeneity parameters in a random coefficients logit model due to this omission and confirmed this result using scanner panel data on margarine purchases. Park and Gupta (2008) pointed out that the omission can cause upward or downward biases in mean and/or heterogeneity parameters. For example, they found strong evidence in a simulation study and in an empirical application to paper towel data that ignoring unmeasured product characteristics may lead to downward biases, regardless of whether these characteristics generate an endogeneity problem or not.

In recent years, researchers have made important advances in choice models by incorporating parameter dynamics. In particular, by introducing parameters that change based on a stochastic process, they have captured changes in consumers' Internet browsing goals (Montgomery et al. 2004), changes in preferences (Kim et al. 2005; Lachaab et al. 2006), and changes in latent relationship states with the firm (Netzer et al. 2007). These advances offer a number of advantages: improved model fits, improved predictions, rich managerial implications, and rigorous understanding of consumer behavior. Kim et al. (2005) and Lachaab et al. (2006) develop flexible time-varying parameter models using a general VAR process within the Bayesian estimation framework. A key feature of the VAR process is that "coefficients for different individuals evolve in the same manner as the population mean" (Lachaab et al. 2006, page 62). That is, their models only account for the dynamics of population parameters (we refer to this as *population-level dynamics*). Montgomery et al. (2004) and Netzer et al. (2007) develop hidden Markov models (or Markov regime-switching models) in which latent states exist and individuals' resident states evolve over time following a Markov process. Key distinctive features of this model, compared to the VAR approach, are: 1) parameters evolve at the individual-level or idiosyncratically (we refer to this as *individual-level dynamics*); and 2) parameters change discretely rather than continuously as in the VAR process[37].

---

[37] Continuous changes in parameters can be approximated in Markov regime-switching models by letting the number of regimes increase. Models that approximate the continuous drift of parameters by a number of discrete regimes may have managerial appeal in many applications. A tractable number of regimes can be easily characterized and interpreted based on the parameter estimates.

**Table 3.3. Comparison of Proposed Model with Other Random Utility Models**

| Study | Heterogeneity | Endogeneity | Parameter Dynamics |
|---|---|---|---|
| Villas-Boas & Winer (1999) | No | **Yes** | No |
| Sudhir (2001) | **Yes** | **Yes** | No |
| Draganska & Jain (2002) | **Yes** | **Yes** | No |
| Yang et al. (2003) | **Yes** | **Yes** | No |
| Goolsbee & Petrin (2004) | **Yes** | **Yes** | No |
| Petrin & Train (2004) | **Yes** | **Yes** | No |
| Montgomery et al. (2004) | **Yes** | No | **Yes (individual-level)** |
| Kim et al. (2005) | **Yes** | No | **Yes (population-level)** |
| Villas-Boas & Zhao (2005) | **Yes** | **Yes** | No |
| Chintagunta et al. (2005) | **Yes** | **Yes** | No |
| Lachaab et al. (2006) | **Yes** | No | **Yes (population-level)** |
| Netzer et al. (2007) | **Yes** | No | **Yes (individual-level)** |
| **This study (2009)** | **Yes** | **Yes** | **Yes (individual-level)** |

In this paper, we propose a unique Markov regime-switching random coefficient logit model that incorporates individual-level dynamics. Along with parameter dynamics, we consider unmeasured product characteristics and the endogeneity of marketing mix variables. The latter has been overlooked in all of the extant choice models with parameter dynamics (Table 3.3 summarizes selected studies). In summary, the proposed model incorporates several important aspects of the consumer choice model - heterogeneity across consumers, endogeneity of marketing mix variables, and parameter dynamics – in as complete a manner as possible.

### 3.3 Model formulation and estimation

### 3.3.1 Model formulation

We assume that on each shopping trip consumers either choose a brand that gives them the highest utility in the category or choose not to purchase in the category. In this paper, we model purchase incidence and brand choice behaviors.[38] On each purchase occasion $t = 1,...,\ T_h$, the utility for the outside good and the utility of brand $j=1,...,J$ in store $s=1,...,S$ for consumer $h=1,...,H$ is given by the following expression:

$$U_{h0t}^s = \alpha_{h,t} + INV_{ht}\beta_{INV,h} + \varepsilon_{h0t}^s, \quad \text{if no purchase}, \tag{1}$$

$$U_{hjt}^s = \beta_{j,h} + \beta_{s,h} + P_{jt}^s\beta_{P,h} + F_{jt}^s\beta_{F,h} + D_{jt}^s\beta_{D,h} + SD_{hjt}\beta_{SD,h} + \xi_{jt}^s + \varepsilon_{hjt}^s, j=1,...,J \tag{2}$$

where $INV_{ht}$ is category inventory carried by household $h$ at beginning of purchase occasion (or shopping trip) $t$, and $P_{jt}^s$, $F_{jt}^s$ and $D_{jt}^s$ are price, feature and display of brand $j$ in store $s$ at $t$, respectively.[39] We incorporate state-dependence through $SD_{hjt}$ which is equal to 1 if the last purchased brand is $j$ and zero, otherwise. $\{\beta_{j,t}\}_{j=1,...,J}$,

---

[38] Purchase quantity is included via an inventory variable that affects the likelihood of buying in the category (i.e. incidence). We do not model the purchase quantity decision. This is reasonable in our empirical application to yogurt, a perishable category in which there is not much quantity variation across occasions (within consumer).

[39] Following Chintagunta et al. (2002), we specify the utility of the outside good as in (1). This specification implies that the preference ordering within the choice set is assumed to be unaffected by the preference orderings in any choice sets that make up the outside good ("weakly separable").

$\{\beta_{s,h}\}_{s=1,...,S}$, $\beta_{P,h}$, $\beta_{F,h}$, and $\beta_{D,h}$ represent individual specific preferences for brands and stores and responsiveness to price, feature, and display, respectively. $\alpha_{h,t}$ represents the utility of the "No purchase" option for consumer $h$ at time $t$. $\beta_{INV,h}$ represents the influence of consumer inventory on category purchase decisions. $\beta_{SD,h}$ captures individual-specific state dependence effect. A positive coefficient implies positive state dependence, or inertia, whereas a negative coefficient implies variety seeking. $\{\varepsilon_{hjt}^s\}_{j=0,1,...,J,s=1,...,S}$ are i.i.d. random shocks with a Type-I Extreme Value distribution.

$\xi_{jt}^s$ are the unmeasured product characteristics (UPC) which may include, for example, the impact of unobserved promotional activity, coupon availability, shelf space, national advertising, unquantifiable factors and systematic shocks to demand. We can raise two issues related to $\xi_{jt}^s$. The first issue is the endogeneity problem. If marketers make their decisions based on the values of $\xi_{jt}^s$, marketing mix variables would be correlated with $\xi_{jt}^s$. In particular, empirical research has typically reported a positive correlation between price and $\xi_{jt}^s$ (or price endogeneity) in disaggregate as well as in aggregate data. Due to this correlation, $\xi_{jt}^s$ is not necessarily mean zero given marketing mix variables and thus, we cannot treat it as another error component and integrate it out of the demand function. Further, regardless of the correlation with marketing mix variables, ignoring $\xi_{jt}^s$ would force the model to absorb these effects in the i.i.d. random shock and/or the remaining explained part of the utility. As a result, one could get biased estimates of model parameters (Chintagunta et al. 2005; Park and Gupta 2008).

A key feature of our model is consumers' switching between high and low category purchase tendencies. To capture this, we use a Markov regime-switching random coefficients framework. The resulting model divides the purchase stream of each consumer into distinct regimes with different parameter values, and regime-

119

switching dynamics follow a first order Markov process. We describe a two-regime

model and also apply this model to data. Extension to three or more regimes is

straight-forward. We specify model parameters as follows:

$$\alpha_{h,t} = \overline{\alpha}_0 (1 - S_{ht}) + \overline{\alpha}_1 S_{ht} + a_h, \quad a_h \sim N(0, \sigma_a^2), \tag{3}$$

$$Prob(S_{ht} = 0 \mid S_{ht-1} = 0) = \frac{\exp(q_1 + q_2 H_h)}{1 + \exp(q_1 + q_2 H_h)} = q_h, \tag{4}$$

$$Prob(S_{ht} = 1 \mid S_{ht-1} = 1) = \frac{\exp(p_1 + p_2 H_h)}{1 + \exp(p_1 + p_2 H_h)} = p_h, \tag{5}$$

$$\beta_h = \overline{\beta} + b_h, \quad b_h \sim N(0, \Sigma), \tag{6}$$

where $\beta_h$ is a vector of $\beta$'s, $\Sigma$ is a covariance matrix of heterogeneity parameters,

and $H_h$ is a household-specific variable which represents household demographics or

characterizes household consumption behaviors. $S_{ht}$ is an indicator of unobservable

discrete states which take values one or zero. Note that the parameter dynamics in

$\alpha_{h,t}$ are dominated by $S_{ht}$ which are individual-specific and evolve idiosyncratically.

These are at the individual-level and model each consumer's switching behavior

between high and low category purchase tendencies. For illustration, let us assume

that $\overline{\alpha}_0 < \overline{\alpha}_1$, $S_{ht} = 0$ for $t=1,...,10$, and $S_{ht} = 1$ for $t=11,...,20$. Then (1), the utility of

"No purchase" becomes $U^s_{h0t} = \overline{\alpha}_0 + a_h + INV_{h,t} \beta_{INV,h} + \varepsilon^s_{h0t}$ for $t=1,...,10$, and

$U^s_{h0t} = \overline{\alpha}_1 + a_h + INV_{h,t} \beta_{INV,h} + \varepsilon^s_{h0t}$ for $t=11,...,20$. Since $\overline{\alpha}_0 < \overline{\alpha}_1$, the probability of

no purchase is lower (i.e., the probability of category purchase is higher) when

$t=1,...,10$ than when $t=11,...,20$. In this manner, we operationalize high and low

purchase tendencies. We expect that the level of inventory in substitute categories is

low (high) during the high (low) purchase tendency period. Thus, $\overline{\alpha}_0$ and $\overline{\alpha}_1$ can be

interpreted as low and high levels of inventory in substitute categories, respectively.

Moreover, we expect that $\alpha_{h,t}$ is negatively correlated with $INV_{h,t}$. $S_{ht}$ follows the

household-specific first order Markov process specified in (4) and (5). $p_1$, $p_2$, $q_1$,

and $q_2$ link *the likelihood* or *the probability* of regime switches to household specific

characteristics, $H_h$. Also, $p_h$ and $q_h$ contain information on the expected duration of

a regime. The expected durations of regimes 0 (where $S_{ht} = 0$) and 1 (where $S_{ht} = 1$) for $h$ can be derived as $1/(1-q_h)$ and $1/(1-p_h)$, respectively.

### 3.3.2 Model estimation

There are three issues in the estimation of the proposed model: the first is how to make inferences on the unobservable discrete state indicator $S_{ht}$, the second is how to handle heterogeneity related to $a$'s and $b$'s, and the third is how to handle the unmeasured product characteristics $\xi_{jt}^s$ and the endogeneity of marketing mix variables.[40] To make inferences on the unobservable discrete state indicator $S_{ht}$, we first consider the joint density of the observed outcome and the state indicator, and then integrate the state indicator out of the joint density by summing over all possible values of the state indicator (Hamilton 1989). To handle heterogeneity, we use the simulated maximum likelihood estimator (SMLE). To handle $\xi_{jt}^s$ and the related endogeneity issue, we use the control function method.

Following Kuksov and Villas-Boas (2007) and Chintagunta et al. (2005), we assume that price is the only endogenous marketing mix variable. Further we assume the following:

$$P_{jt}^s = Z_{jt}'^s \gamma_j^s + v_{jt}^s, \quad v_{jt}^s \sim i.i.d. \, N(0, \sigma_{v_{js}}^2), \tag{7}$$

$$\xi_{jt}^s \sim i.i.d. \, N(0, \sigma_{\xi_{js}}^2), \tag{8}$$

$$Cov(v_{jt}^s, \xi_{jt}^s) = \lambda_{js}, \tag{9}$$

$$Cov(Z_{jt}^s, \xi_{jt}^s) = 0, \tag{10}$$

where $Z_{jt}^s$ is an *L*-dimensional vector of instrumental variables uncorrelated with $\xi_{jt}^s$ but correlated with $P_{jt}^s$. We apply the Cholesky decomposition of the covariance matrix of $[v_{jt}^s \; \xi_{jt}^s]'$ in order to rewrite it as a function of two independent shocks:

---

[40] Since we observe a household's decision only when it visits stores, *t* denotes purchase occasion in (1) – (6). However, when we use *t* in terms which are common to all households, such as $\xi_{jt}^s$, *t* denotes calendar time.

$$\begin{bmatrix} v_{jt}^s \\ \xi_{jt}^s \end{bmatrix} = \begin{bmatrix} \theta_{11,j}^s & 0 \\ \theta_{21,j}^s & \theta_{22,j}^s \end{bmatrix} \begin{bmatrix} \omega_{1,jt}^s \\ \omega_{2,jt}^s \end{bmatrix}, \quad \begin{bmatrix} \omega_{1,jt}^s \\ \omega_{2,jt}^s \end{bmatrix} \sim i.i.d.N\left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right), \tag{11}$$

where $\theta_{11,j}^s = \sigma_{v_{js}}$. From (11), we can rewrite $\xi_{jt}^s$ as follows:

$$\begin{aligned} \xi_{jt}^s &= \theta_{21,j}^s (\theta_{11,j}^s)^{-1} v_{jt}^s + \theta_{22,j}^s \omega_{2,jt}^s \\ &= \theta_{21,j}^s (\theta_{11,j}^s)^{-1} \cdot (P_{jt}^s - Z_{jt}^s \gamma_j^s) + \theta_{22,j}^s \omega_{2,jt}^s \\ &= \theta_j^s \cdot CF_{jt}^s + \varsigma_{jt}^s \end{aligned} \tag{12}$$

where $\theta_j^s = \theta_{21,j}^s (\theta_{11,j}^s)^{-1}$ and $CF_{jt}^s = P_{jt}^s - Z_{jt}^s \gamma_j^s$. Using (12), the utility specification of (2) becomes,

$$U_{hjt}^s = \beta_{j,h} + \beta_{s,h} + P_{jt}^s \beta_{P,h} + F_{jt}^s \beta_{F,h} + D_{jt}^s \beta_{D,h} + SD_{hjt} \beta_{SD,h} + \theta_j^s CF_{jt}^s + \varsigma_{jt}^s + \varepsilon_{hjt}^s. \tag{13}$$

Note that after introducing $CF_{jt}^s$, the new error term $\varsigma_{jt}^s$ is uncorrelated with any other term in (13). We refer to this $\varsigma_{jt}^s$ as the exogenous unmeasured product characteristics (EUPCs). While EUPCs do not generate any endogeneity issue, they represent the exogenous common shocks which might cause biases if ignored (Chintagunta et al. 2005; Park and Gupta 2008).

One way to estimate (13) is a fixed-effect approach in two stages (Berry 1994; Chintagunta et al. 2005). We can decompose $U_{hjt}^s$ into three parts: the common utility $\delta_{jt}^s$, individual–specific utility $\mu_{hjt}^s$, and Type-I Extreme Value error term $\varepsilon_{hjt}^s$. The common utility $\delta_{jt}^s$ is estimated as a parameter in the first stage along with heterogeneity parameters in $\mu_{hjt}^s$. Then, the parameters of interest (i.e. $\bar{\beta}$) are recovered from the estimate of $\delta_{jt}^s$ in the second stage. This approach is easy to understand and straight-forward to implement. However, certain difficulties arise which might preclude researchers from using this method. First, the number of parameters to be estimated in the first stage is large. If the data come from a single store, the number of parameters for the common utility is $(J \times T)^{41}$. We use

---

[41] $T$ denotes a number of elements in the superset of calendar times in $\{T_h\}_{h=1,\dots,H}$.

household panel data which come from $S$ stores and thus we have $(J \times T \times S)$ parameters for the common utility as well as heterogeneity parameters (a total of 1,244 parameters in our case). Considering the fact that researchers usually need numerical integration to handle heterogeneity in the first-stage estimation, this method is computationally very expensive, sometimes infeasible. Another challenge with the fixed-effect approach arises due to its data requirements: it requires more than one choice observation for each alternative at each store in each of the time periods under consideration. Without such information, it is difficult to identify alternative-, store-, and time-specific common utility since there is no information on alternatives at the store during the periods in which they are not purchased by any of the panel households (Chintagunta and Dubé 2005). In the empirical analysis in Section 3.4, we use data from 519 panel households but do not have one or more purchases for each alternative at each store in each time period. Chintagunta and Dubé (2005) report the same problem when using household panel data at the SKU level.

To circumvent the above-mentioned difficulties, we propose an approximation method. Specifically, we approximate the EUPC $\varsigma_{jt}^{s}$ by using $\varsigma_{jm} + \varsigma_{s}$ ($m = 1,...,M$; $s = 1,...,S$). $\varsigma_{s}$ represents store-specific mean level of EUPC. Here we assume that the store-specific factors in EUPC are independent of brand- and time-specific factors in EUPC. $m$ represents a time period which is longer than the time period represented by $t$. For example, if $t$ represents week, then $m$ can be a month or a quarter. While we are not able to observe one or more purchases for all of $(J \times T \times S)$, we might be able to observe one or more purchases for all of $(J \times T \times M)$. Since we cannot separately identify $\{\varsigma_{jm}\}_{j=1,..,J,m=1,...,M}$, $\{\varsigma_{s}\}_{s=1,..,S}$, $\{\overline{\beta}_{j}\}_{j=1,...,J}$, and $\{\overline{\beta}_{s}\}_{s=1,...,S}$, we take a two-stage estimation approach. In the first stage, we estimate $\{\zeta_{jm}\}_{j=1,..,J,m=1,...,M}$ and $\{\zeta_{s}\}_{s=1,...,S}$ where $\zeta_{jm} = \varsigma_{jm} + \overline{\beta}_{j}$ and $\zeta_{s} = \varsigma_{s} + \overline{\beta}_{s}$ along with other parameters. Then,

we estimate $\{\bar{\beta}_j\}_{j=1,...,J}$ from the estimates of $\zeta_{jm}$'s in the second stage.[42] Note that this method approximates only the exogenous part of $\xi_{jt}^s$ and the endogeneity problem between $\xi_{jt}^s$ and price is handled exactly through $CF_{jt}^s$.

To calculate household inventory $INV_{ht}$, we use the following inventory identity (Bucklin and Lattin 1991; Gupta 1988; Ailawadi and Neslin 1998):

$$INV_{ht} = INV_{ht-1} + PurQty_{ht-1} - Consumpt_{ht-1},$$

where $PurQty_{ht-1}$ is quantity purchased by household $h$ at purchase occasion $t$-$1$, and $Consumpt_{ht-1}$ is consumption by household $h$ since $t$-$1$. Because consumption data are not available, $Consumpt_{ht-1}$ is calculated from the average consumption rate of household $h$ which is estimated as follows:

$$AveConsumptRate_h = \Sigma_{t'}PurQty_{ht'} / \Sigma_{t'} \min\{10\,weeks, InterPurchaseTime_{ht'}\},$$

where $t'$ indexes store visits on which the category was purchased, and $InterPurchaseTime_{ht'}$ denotes time elapsed since last category purchase. This calculation is based on the following assumptions: if $InterPurchaseTime_{ht'}$ is less than 10 weeks, a household is assumed to consume $PurQty_{ht-1}$ at a constant rate over $InterPurchaseTime_{ht'}$; if $InterPurchaseTime_{ht'}$ is equal to or greater than 10 weeks, a household is assumed to consume $PurQty_{ht-1}$ at a constant rate over 10 weeks. We consider the 10-week threshold because of perishability of yogurt.[43] In the absence of the 10-week threshold, our estimate of average consumption rate is equivalent to the consumption

---

[42] Since we use $\hat{\zeta}_{jm}$ instead of $\zeta_{jm}$, the standard errors of the second stage estimates from the traditional formulas are biased downward. To approximate this additional source of variance, we bootstrap the second stage estimation. Specifically, we resample $\hat{\zeta}_{jm}$ from its asymptotic distribution identified in the first stage and re-estimate the second stage estimates. We repeat this exercise over many re-sampled $\hat{\zeta}_{jm}$ and the variance in the parameter estimates across the bootstrapped sample is then added to the variance from the traditional formulas.

[43] We also tried an 8-week upper limit, instead of 10-week. Another estimate of the average consumption rate we tried is the starting inventory, as in Ailawadi and Neslin (1998). The empirical results shown in Section 4 were not substantially affected.

rate of Gupta (1988). Also, the starting inventory for each household is set equal to the average purchase quantity (Gupta 1988).[44]

Now, we discuss the details of the handling of heterogeneity and unobservable discrete state indicator $S_{ht}$. Let $y_{ht}^s = i$ denote the event where consumer $h$ chooses brand $i$ in store $s$ at time $t$ and $y_{ht}^s = 0$ denotes no category purchase. $\psi_{ht}$ denotes a set of information about household $h$ up to time $t$. $i_{ht}^s$ denotes the chosen alternative of household $h$ in store $s$ at time $t$ and $i_h = \{i_{ht}^s\}_{t=1,...,T_h, s=1,...,S}$ denotes a sequence of choices made by household $h$. For expositional convenience, we first assume that heterogeneity parameter $\Delta_h = \{a_h, b_h\}$ are given. The probability that household $h$ makes the sequence of choices $i_h$ is

$$L_h(\Delta_h) = \prod_{t=1}^{T_h} f(y_{ht}^s = i_{ht}^s \mid \psi_{ht-1}) = \prod_{t=1}^{T_h} \sum_{S_{ht}=0}^{1} f(y_{ht}^s = i_{ht}^s, S_{ht} \mid \psi_{ht-1})$$

$$= \prod_{t=1}^{T_h} \sum_{S_{ht}=0}^{1} f(y_{ht}^s = i_{ht}^s \mid \psi_{ht-1}, S_{ht}) \times P(S_{ht} \mid \psi_{ht-1}). \tag{14}$$

Conditional on $\Delta_h$ and $S_{ht}$, $f(y_{ht}^s = i_{ht}^s \mid \psi_{ht-1}, S_{ht})$ is the standard logit probability, since $\varepsilon$'s have iid Type-I Extreme Value distributions.

$$f(y_{ht}^s = i_{ht}^s \mid \psi_{ht-1}, S_{ht} = 0) = \begin{cases} \dfrac{\exp(\overline{\alpha}_0 + a_h + INV_{ht}\beta_{INV,h})}{\exp(\overline{\alpha}_0 + a_h + INV_{ht}\beta_{INV,h}) + \sum_{k=1}^{J} \exp(V_{hkt}^s)}, & i_{ht}^s = 0 \\[2em] \dfrac{\exp(V_{hi_{ht}^s t}^s)}{\exp(\overline{\alpha}_0 + a_h + INV_{ht}\beta_{INV,h}) + \sum_{k=1}^{J} \exp(V_{hkt}^s)}, & i_{ht}^s \neq 0 \end{cases}$$

$$\tag{15}$$

---

[44] Some previous studies (e.g. Erdem et al. 2003) have noted that the constructed inventory variables may cause problems in the analysis. Consequently, we used three variables instead of constructed inventory variables: 1) lag purchase quantity, 2) time elapsed since last purchase, and 3) interaction of 1) and 2). These three variables act as proxy variables or instrumental variables for unobservable inventory. The results from this method were not different from the results based on the constructed inventory variables. Results are available from the authors on request.

$$
f(y_{ht}^s = i_{ht}^s \mid \psi_{ht-1}, S_{ht} = 1) = \begin{cases} \dfrac{\exp(\overline{\alpha}_1 + a_h + INV_{ht}\beta_{INV,h})}{\exp(\overline{\alpha}_1 + a_h + INV_{ht}\beta_{INV,h}) + \sum_{k=1}^{J}\exp(V_{hkt}^s)}, & i_{ht}^s = 0 \\[4mm] \dfrac{\exp(V_{hi_{ht}^s t}^s)}{\exp(\overline{\alpha}_1 + a_h + INV_{ht}\beta_{INV,h}) + \sum_{k=1}^{J}\exp(V_{hkt}^s)}, & i_{ht}^s \neq 0 \end{cases}
$$
$$(16)$$

where $V_{hjt}^s = \beta_{j,h} + \beta_{s,h} + P_{jt}^s\beta_{P,h} + F_{jt}^s\beta_{F,h} + D_{jt}^s\beta_{D,h} + SD_{hjt}\beta_{SD,h} + \theta_j^s CF_{jt}^s + \varsigma_{mt}^s$.

$P(S_{ht} \mid \psi_{ht-1})$ can be decomposed as follows:

$$
\begin{aligned}
P(S_{ht} \mid \psi_{ht-1}) &= \sum_{S_{ht-1}=0}^{1} P(S_{ht}, S_{ht-1} \mid \psi_{ht-1}) \\[2mm]
&= \sum_{S_{ht-1}=0}^{1} P(S_{ht} \mid S_{ht-1}, \psi_{ht-1}) \times P(S_{ht-1} \mid \psi_{ht-1}) \\[2mm]
&= \sum_{S_{ht-1}=0}^{1} P(S_{ht} \mid S_{ht-1}) \times P(S_{ht-1} \mid \psi_{ht-1})
\end{aligned}
$$
$$(17)$$

where $P(S_{ht} \mid S_{ht-1})$ is determined by transition probabilities $p_h$ and $q_h$. Given $P(S_{ht-1} \mid \psi_{ht-1})$, we can calculate $f(y_{ht}^s = i_{ht}^s \mid \psi_{ht-1})$ using (14) - (17). For $f(y_{ht+1}^s = i_{ht+1}^s \mid \psi_{ht})$, we need $P(S_{ht} \mid \psi_{ht})$, which can be updated as follows:

$$
\begin{aligned}
P(S_{ht} \mid \psi_{ht}) = P(S_{ht} \mid \psi_{ht-1}, y_{ht}^s = i_{ht}^s) &= \frac{f(S_{ht}, y_{ht}^s = i_{ht}^s \mid \psi_{ht-1})}{f(y_{ht}^s = i_{ht}^s \mid \psi_{ht-1})} \\[2mm]
&= \frac{f(y_{ht}^s = i_{ht}^s \mid S_{ht}, \psi_{ht-1}) P(S_{ht} \mid \psi_{ht-1})}{f(y_{ht}^s = i_{ht}^s \mid \psi_{ht-1})}
\end{aligned}
$$
$$(18)$$

Now we can calculate $f(y_{ht}^s = i_{ht}^s \mid \psi_{ht-1})$ for all $t$ given $P(S_{h0} \mid \psi_{h0})$. We can use the steady-state probabilities for $P(S_{h0} \mid \psi_{h0})$,

$$
P(S_{h0} = 0 \mid \psi_{h0}) = \frac{1 - p_h}{1 - p_h - q_h}
$$
$$(19)$$

$$
P(S_{h0} = 1 \mid \psi_{h0}) = \frac{1 - q_h}{1 - p_h - q_h}
$$
$$(20)$$

The probability that household $h$ makes the sequence of choices $i_h$ is nothing but the product of $f(y_{ht}^s = i_{ht}^s \mid \psi_{ht-1})$'s for all $t$.

So far, we have assumed that $\Delta_h$ is given. However, we do not know $\Delta_h$ and therefore we do not know $L_h(\Delta_h)$, which is conditional on $\Delta_h$. The unconditional probability is the integral of $L_h(\Delta_h)$ over all possible values of $\Delta_h$:

$$L_h = \int L_h(\Delta_h)\, \phi(\Delta_h)\, d\Delta_h \tag{21}$$

where $\phi(\cdot)$ denotes a multivariate normal density of $\Delta_h$. By maximizing $\Sigma_h \ln L_h$, we get the estimates of parameters. The integral in $L_h$ is approximated through Monte Carlo simulation methods and thus, the resulting estimate is a Simulated Maximum Likelihood Estimator (see Keane 1993).

To provide an intuitive explanation, the identification of the model comes from the partitioning of panel households' observations into different regimes, creating "pseudo" households with different preference parameters that are identified using long purchase histories available for each household. To formally show the model identification, we performed a simulation study that is discussed in Appendix 2.

### 3.3.3 Individual-level parameters and smoothed transition probability

Individual-level parameters are valuable since a goal in the analysis of household choice data is not only to describe the extent and nature of household heterogeneity but also to make inferences about specific households for customizing various marketing actions (Allenby and Rossi 1999). We use Bayes rule to calculate expected individual-level parameters $\Delta_h$ conditional on the individual's purchase history and the population-level parameter estimate $\hat{\Theta}$:

$$\hat{\Delta}_h = \frac{\int \Delta_h L_h(\Delta_h)\, \phi(\Delta_h)\, d\Delta_h}{\int L_h(\Delta_h)\, \phi(\Delta_h)\, d\Delta_h}. \tag{22}$$

The major weakness of this approach is that it does not account for the uncertainty in individual-level parameters due to estimation error in the population-level parameter $\Theta$. Following Revelt and Train (2000), we apply a bootstrap method to overcome this weakness. Specifically, we resample the population-level parameters from their asymptotic distributions and calculate (22) using each drawn parameter. Given $\hat{\Theta}$ and $\hat{\Delta}_h$, we can make inferences on $S_{ht}$ using all of the information in the sample. This gives us the smoothed probability $\{P(S_{ht} | \psi_{hT_h})\}_{t=1}^{T_h}$.[45] For each re-sampled parameter, we calculate the smoothed probability $\{P(S_{ht} | \psi_{hT_h})\}_{t=1}^{T_h}$. By repeating this procedure many times, we get the confidence intervals of $S_{ht}$.

### 3.3.4 Benchmark Models

In addition to the proposed model, we estimate three benchmark models[46] that allow us to do the following: 1) compare the performance of the proposed model to that of a representative extant approach that accounts for serial dependence in category purchase decision; and 2) decide whether the sample yogurt data we use in our empirical analysis can be better characterized by switching behavior in the category purchase decision, or by the alternation between inertial and variety seeking states in the brand choice decision.

**Benchmark 1**

$$U_{h0t}^s = \beta_{0,h} + INV_{ht}\beta_{INV,h} + LagInc_{ht}\beta_{LagInc,h} + \varepsilon_{h0t}^s, \quad \text{if no purchase,} \tag{23}$$

$$U_{hjt}^s = \beta_{j,h} + \beta_{s,h} + P_{jt}^s\beta_{P,h} + F_{jt}^s\beta_{F,h} + D_{jt}^s\beta_{D,h} + SD_{hjt}\beta_{SD,h} + \xi_{jt}^s + \varepsilon_{hjt}^s, \quad j=1,\dots,J \tag{24}$$

---

[45] Smoothed probability is different from filtered probability $P(S_{ht} | \psi_{ht})$. The former uses all the information in the sample and the latter uses the information up to $t$ to make inferences on $S_{ht}$. In this paper, we use the smoothing algorithm proposed by Kim (1994). This algorithm is more efficient than the one proposed in Hamilton (1989).

[46] As alternatives to random utility models, marketing researchers also have employed probability models such as Markov Chains to study consumers' purchase behavior. Here, all our benchmark models are based on the random utility approach. We justify our choice of benchmark models based on the following considerations: 1) it is reported that both models (i.e. a random utility model and a probability model) are remarkably similar in terms of both prediction and recovery of marketing-mix elasticities (Seetharaman 2003); 2) a random utility model enables price endogeneity to be accommodated, which is one of the contributions of the proposed model.

In this model, a lagged incidence variable $LagInc_{ht}$ is added to the usual random coefficient logit model in order to capture serial dependence in category purchase incidence (Ailawadi and Neslin 1998). Since our main interest is in the serial dependence in category purchase incidence, it is reasonable to consider this model as a starting benchmark. $LagInc_{ht}$ is equal to 1 if yogurt was purchased on the last shopping trip and zero otherwise. This captures state-dependence in the category purchase. Note that state-dependence in brand choice is separately considered by the term $SD_{hjt}$. Also, we consider the unmeasured product characteristics and price endogeneity. To estimate this model, we use the control function approach with the approximation of EUPC.

As noted previously, Markov regime-switching coefficient $\alpha_{h,t}$ captures unobservable inventory in substitute categories, which is expected to be negatively correlated with $INV_{ht}$. If the substitute inventory is an important factor in a consumer's utility but cannot appropriately be explained by $LagInc_{ht}$, then $\varepsilon_{hjt}^{s}$ should absorb the influence of the substitute inventory. Consequently, an endogeneity problem arises in (23) (we refer to this as an "inventory endogeneity problem" hereafter) due to the negative correlation between the regressor $INV_{ht}$ and the random error $\varepsilon_{hjt}^{s}$. Moreover, we can expect that this endogeneity problem will bias the estimate of $\beta_{INV,h}$ negatively. In the next section, we show that the inventory endogeneity problem does arise in our data. Also, we show that we can circumvent this problem by introducing Markov regime-switching coefficient $\alpha_{h,t}$ as in the proposed model.

**Benchmark 2**

$$U_{h0t}^{s} = \beta_{0,h} + INV_{ht}\beta_{INV,h} + TE_{ht}\beta_{TE1,h} + TE_{ht}^{2}\beta_{TE2,h} + (1/TE_{ht})\beta_{TE3,h} + \varepsilon_{h0t}^{s} \quad (25)$$

$$U_{hjt}^{s} = \beta_{j,h} + \beta_{s,h} + P_{jt}^{s}\beta_{P,h} + F_{jt}^{s}\beta_{F,h} + D_{jt}^{s}\beta_{D,h} + SD_{hjt}\beta_{SD,h} + \xi_{jt}^{s} + \varepsilon_{hjt}^{s}, j=1,...,J \quad (26)$$

The difference relative to Benchmark 1 is the inclusion in $U_{h0t}^s$ of three functions of $TE_{ht}$, which is defined as Time Elapsed since the last purchase, instead of $LagInc_{ht}$. The idea is that $LagInc_{ht}$ might not be flexible enough to model serial dependence in category purchase incidence, and the use of three functions of $TE_{ht}$ offers greater flexibility in this regard. This specification allows for a wide range of hazard shapes and has therefore been used successfully in modeling of inter-purchase time (Singh et al. 2006). Also, we can regard these three functions as proxies for substitute inventory which may be related with $TE_{ht}$. If so, the inventory endogeneity problem will also be remedied.

**Benchmark 3**

$$U_{h0t}^s = \beta_{0,h} + INV_{ht}\beta_{INV,h} + \varepsilon_{h0t}^s, \qquad \text{if no purchase,} \qquad (27)$$

$$U_{hjt}^s = \beta_{j,h} + \beta_{s,h} + P_{jt}^s\beta_{P,h} + F_{jt}^s\beta_{F,h} + D_{jt}^s\beta_{D,h} + SD_{hjt}\alpha_{SD,ht} + \xi_{jt}^s + \varepsilon_{hjt}^s, \ j=1,...,J \ (28)$$

In this model, we consider the unmeasured product characteristics, price endogeneity, and parameter dynamics. The difference relative to the proposed model is that this benchmark model incorporates individual-level parameter dynamics in the coefficient of the state-dependence term $SD_{hjt}$:

$$\alpha_{SD,ht} = \overline{\alpha}_{SD,0}(1 - S_{\alpha,SD,ht}) + \overline{\alpha}_{SD,1}S_{\alpha,SD,ht} + a_h \qquad (29)$$

By comparing this benchmark to the proposed model in the data, we can test whether the switching behavior in consumers' yogurt purchase can be better characterized in the category purchase decision or in the brand choice decision.

**3.4 Empirical analysis**

**3.4.1 Data**

The data used in the study are histories of yogurt purchases of IRI BehaviorScan panel households (HHs) in Eau Claire, Wisconsin over 104 weeks in 2003 and 2004. The yogurt category is appropriate for the application of the proposed model for at least two reasons. First, yogurt can be categorized as a hedonic good and

the behavioral factors mentioned in section 3.1 (i.e. need for stimulation, satisfaction, boredom, satiation, or seeking balance) are potentially relevant in explaining the consumption of yogurt. Second, yogurt is highly perishable and therefore offers low opportunity to stockpile. Hence, a consumer's switching behavior in category purchase is mostly due to the consumer's underlying consumption behavior, not due to stockpiling.

Among 2,582 panel HHs in this market, 2,206 HHs purchased yogurt (excluding drinking yogurt) more than once during the sample period. We focus on the three largest grocery stores (IRI key: 233779, 264075, and 653776). Purchases at these three stores amount to 80% of the total yogurt purchases in this area. Out of 899 panelist HHs who purchased yogurt only in these three stores, we use data of 519 HHs (58%) who purchased yogurt seven or more times during the 104-week sample period. The use of frequent buyers helps the model identify switching behavior between high and low purchase tendencies. Notably, the selected 519 HHs account for 92% of total yogurt purchases made by the 899 HHs, thus constituting a segment that is of interest to firms in this industry. Hence, we believe our selection rule does not compromise sample representativeness too much.

We include the three largest yogurt brands (Yoplait, Dannon, and Wells) as well as "Others" in the analysis. Sales of these three major brands account for 77% of total category sales in our sample. Additionally, we include a "No purchase" option defined as shopping visits without yogurt purchase. Within each brand, the purchase of any one of the different package sizes was counted as a purchase of the brand.

**Table 3.4: Descriptive statistics of Yogurt Data**

(Number of households = 519, Number of weeks = 104, Number of trips = 86,492)

| *Brands* | *Number of purchases* | *Price ($ per pt.)* | | *Feature (% of store weeks)* | *Display (% of store weeks)* |
| --- | --- | --- | --- | --- | --- |
| | | Mean | SD | | |
| Yoplait | 5,037 | 1.68 | 0.14 | 32% | 19% |
| Dannon | 2,972 | 1.50 | 0.12 | 21% | 15% |
| Wells | 1,723 | 1.37 | 0.10 | 12% | 10% |
| Others | 3,362 | 1.46 | 0.15 | 28% | 28% |

Prices are defined on a per pint basis in our analysis. Price, display, and feature at the brand level were computed as market share-weighted averages of UPC-level variables. Descriptive statistics of the purchases and marketing mix variables are in Table 3.4. About 15% of store visits result in purchases of yogurt. Yoplait is the dominant brand in the market, with 40% market share.

To control for the endogeneity of price, we use prices at the other stores and quarter dummies as instruments. The price of a brand at a store is highly correlated with prices of the same brand at other stores because competing retailers are likely to be offered the same wholesale price.[47] However, we do not expect the unmeasured product characteristics, especially those determined at retail (e.g. shelf space allocation), to be systematically related with wholesale prices. To the extent that this expectation is true, our instrumental variables are valid for controlling for the endogeneity of prices.

### 3.4.2 Estimation and results

Table 3.5 reports the estimation results of the proposed model and the three benchmark models. As mentioned previously, all benchmark models and the proposed model are estimated with the control functions and the approximation of the EUPC using the two-stage estimation approach explained in 3.2. In particular, we approximated $\varsigma_{jt}^{s}$ at the quarter level. The estimates of $\{\zeta_{jm}\}_{j=1,..,J,m=1,...,M}$ at the first stage are all significant at the 0.05 level in all models. To conserve space, we do not report them here. For the normalization, we let $\overline{\alpha}_0$, $\beta_{0,h}$, and $\varsigma_{s=1} + \beta_{s=1,h}$ be equal to zero. We specify the covariance matrix of heterogeneity distribution $\Sigma$ as a diagonal matrix. In (4) – (5), we incorporate the household-specific explanatory variable $H_h$ into the transition probabilities. In the estimation of the proposed model, we

---

[47] The R-square of preliminary regressions of prices on instruments is 0.62 on average (max: 0.85, min: 0.32).

considered 1) household size, 2) average purchase quantity, and 3) the number of

yogurt purchases during the sample period as candidates for $H_h$ in the proposed

model. We use the number of yogurt purchases during the sample period since it

shows the best result in model fit measures (i.e. log-likelihood, AIC, and BIC).

Benchmark 1 is the usual random coefficient logit model with a lagged

incidence indicator variable. The mean of the lagged incidence parameter is

significant and negative. This tells us that, on average, category purchase on one

shopping trip is negatively related to the likelihood of "No Purchase," or positively

related with the likelihood of purchase, on the next trip. This supports the empirical

findings of Ailawadi and Neslin (1998) who attribute a positive effect of lagged

incidence on subsequent category purchase to represent "eating bouts, binging, special

diets" etc. Both the mean and the heterogeneity parameters of state-dependence are

positive and significant and they imply that most households are inertial in their brand

choices. As expected, price has a significant negative coefficient and the effects of

feature and display are positive and significant. On average, households' intrinsic

brand preferences can be ordered as $\text{Yoplait} > \text{Dannon} \geq \text{Others} > \text{Wells}$. This

preference order is preserved in all four models. The mean parameter of inventory is

negative and significant and this implies that the more inventory a household has, the

lower the probability of "No purchase." This result is both counterintuitive and the

opposite of what previous research has documented.[48] The same result also occurs in

all three benchmark models. We suspect that this is due to the inventory endogeneity

problem. None of the benchmark models consider inventory of substitute categories,

which is likely negatively correlated with $INV_{ht}$. As a consequence, $\varepsilon_{hjt}^s$ contains the

influence of omitted substitute inventory and a negative correlation between $INV_{ht}$

---

[48]For example, Gupta (1988) and Ailawadi and Neslin (1998) report that the effect of household inventory on the probability of "No Purchase" is positive.

and $\varepsilon_{bjt}^{s}$ biased the estimate of $\beta_{INV,b}$ negatively. The estimated coefficients of 11 of

12 control functions are significant and negative. This result implies that there exists a

significant price endogeneity problem in the data which, if ignored, can cause bias in

the estimated price effect. In the literature, the unmeasured product characteristics are

usually believed to be positively correlated with price, which is consistent with the

commonly reported downward bias in the price coefficient when this correlation is

ignored. However, our result implies negative correlation between prices and

unmeasured product characteristics, one explanation for which lies in expanded shelf

space allocation or favorable shelf locations of price-promoted products, activities that

are unobserved in our data. If shelf space changes are a dominant component of the

unmeasured product characteristics, we can expect these characteristics to be

negatively correlated with prices.

　　　We now consider the effect of including flexible functions of time to capture

the serial dependence in purchase incidences by comparing the results of Benchmarks

1 and 2. In Benchmark 2, we use three functions of $TE_{bt}$ instead of $LagInc_{bt}$ and all

three fit measures (log-likelihood, AIC, and BIC) improve slightly over Benchmark 1.

The mean of estimates for $(TE_{bt}/10)^{2}$ is significant and positive but those of $TE_{bt}/10$

and $1/TE_{bt}$ are not significant. Note that the mean parameter of $INV_{bt}$ is still

negative and significant.

## Table 3.5: Estimation Results of Yogurt Data

(estimates in bold are significant at p = 0.05)

| | *Benchmark 1* | | | | | *Benchmark 2* | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Mean | | Heterogeneity | | | Mean | | Heterogeneity | |
| Variable | Est. | SE | Est. | SE | Variable | Est. | SE | Est. | SE |
| INV | **-0.03** | **0.00** | **0.01** | **0.00** | INV | **-0.07** | **0.01** | **0.02** | **0.00** |
| Yoplait | **-2.26** | **0.13** | **0.50** | **0.03** | Yoplait | **-2.13** | **0.13** | **0.80** | **0.03** |
| Dannon | **-2.76** | **0.13** | **0.36** | **0.03** | Dannon | **-2.65** | **0.13** | **0.77** | **0.03** |
| Wells | **-3.78** | **0.13** | **1.29** | **0.04** | Wells | **-3.56** | **0.13** | **1.59** | **0.05** |
| Others | **-2.76** | **0.13** | **0.47** | **0.04** | Others | **-2.65** | **0.13** | **0.63** | **0.03** |
| Store2 | **0.22** | **0.04** | **0.49** | **0.03** | Store2 | 0.03 | 0.04 | **0.09** | **0.04** |
| Store3 | **-0.13** | **0.04** | **0.18** | **0.03** | Store3 | **0.15** | **0.04** | **0.15** | **0.06** |
| Price | **-0.72** | **0.10** | **0.54** | **0.01** | Price | **-0.69** | **0.10** | **0.47** | **0.01** |
| Feature | **0.28** | **0.05** | **0.19** | **0.07** | Feature | **0.24** | **0.05** | **0.21** | **0.05** |
| Display | **0.99** | **0.07** | **0.20** | **0.09** | Display | **1.00** | **0.07** | **0.25** | **0.09** |
| SD | **0.93** | **0.03** | **0.69** | **0.02** | SD | **0.80** | **0.02** | **0.07** | **0.02** |
| LagInc | **-0.24** | **0.03** | **0.17** | **0.03** | TE/10 | 0.00 | 0.13 | **0.08** | **0.02** |
| | | | | | (TE/10)^2 | **0.18** | **0.02** | **0.04** | **0.01** |
| | | | | | 1/TE | 0.00 | 0.00 | 0.00 | 0.00 |
| 12 CFs | -10.27 to -1.23, all significant at p=0.05 except $CF_{j=1,t}^{s=1}$ | | | | 12 CFs | -10.25 to -1.03, all significant at p=0.05 except $CF_{j=1,t}^{s=1}$ | | | |
| | | | | | | | | - | |
| | LL | | -46341 | | | LL | | 46297 | |
| | AIC | | 92810 | | | AIC | | 92731 | |
| | BIC | | 93410 | | | BIC | | 93368 | |
| | # Params | | 64 | | | # Params | | 68 | |

136

**Table 3.5 (Continued)**

| | *Benchmark 3* | | | | | *Proposed Model* | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Mean | | Heterogeneity | | | Mean | | Heterogeneity | |
| Variable | Est. | SE | Est. | SE | Variable | Est. | SE | Est. | SE |
| INV | **-0.02** | **0.00** | **0.02** | **0.00** | INV | **0.02** | **0.00** | **0.03** | **0.00** |
| Yoplait | **-2.25** | **0.13** | **0.52** | **0.04** | Yoplait | **-0.95** | **0.13** | **1.03** | **0.04** |
| Dannon | **-2.71** | **0.13** | **0.74** | **0.04** | Dannon | **-1.58** | **0.13** | **1.15** | **0.04** |
| Wells | **-3.59** | **0.13** | **1.08** | **0.04** | Wells | **-2.85** | **0.13** | **1.81** | **0.05** |
| Others | **-2.71** | **0.13** | **0.86** | **0.03** | Others | **-1.52** | **0.13** | **1.09** | **0.04** |
| Store2 | **0.32** | **0.04** | **0.27** | **0.05** | Store2 | **0.09** | **0.04** | **0.54** | **0.03** |
| Store3 | **0.27** | **0.04** | **0.39** | **0.04** | Store3 | -0.07 | 0.05 | **0.42** | **0.04** |
| Price | **-0.65** | **0.10** | **0.17** | **0.02** | Price | **-0.64** | **0.10** | **0.16** | **0.02** |
| Feature | **0.30** | **0.05** | **0.15** | **0.05** | Feature | **0.28** | **0.06** | **0.46** | **0.06** |
| Display | **1.08** | **0.07** | 0.10 | 0.08 | Display | **1.07** | **0.07** | **0.35** | **0.11** |
| $\overline{\alpha}_{SD,0}$ | **2.17** | **0.04** | **0.24** | **0.03** | SD | **0.43** | **0.02** | **0.15** | **0.03** |
| $\overline{\alpha}_{SD,1}$ | **0.12** | **0.06** | | | $\overline{\alpha}_1$ | **2.15** | **0.04** | **0.17** | **0.04** |
| Q | **0.99** | **0.01** | | | q1 | **1.87** | **0.22** | | |
| P | **0.99** | **0.01** | | | q2 | **0.78** | **0.09** | | |
| | | | | | p1 | **4.90** | **0.19** | | |
| | | | | | p2 | 0.04 | 0.07 | | |
| 12 CFs | -10.04 to -1.38, all significant at p=0.05 except $CF_{j=1,t}^{s=1}$ | | | | 12 CFs | -9.19 to -0.52, all significant at p=0.05 except $CF_{j=1,t}^{s=1}$ & $CF_{j=3,t}^{s=3}$ | | | |
| | | | | | | - | | | |
| | LL | | -45575 | | | LL | | 43880 | |
| | AIC | | 91278 | | | AIC | | 87897 | |
| | BIC | | 91878 | | | BIC | | 88534 | |
| | # Params | | 64 | | | # Params | | 68 | |

This result tells us that the functions of $TE_{ht}$ do not work as proxies for the substitute inventory. To summarize, incorporating flexible functions of $TE_{ht}$ enables Benchmark 2 to capture serial dependence in purchase incidence better than Benchmark 1 but this does not solve the problem of inventory endogeneity.

Benchmark 3 and the proposed model consider individual-level parameter dynamics as well as unmeasured product characteristics. Compared to Benchmarks 1 and 2, both models are significantly better in all three fit measures, namely, log-likelihood, AIC, and BIC. This supports the existence of switching behavior in consumers' decision making. By comparing Benchmark 3 and the proposed model, we can observe which individual-level dynamics are better in explaining consumers' switching behavior. The estimation results show that the proposed model is significantly better than Benchmark 3 in terms of log-likelihood, AIC, and BIC.[49] Therefore, we conclude that the sample yogurt data can be better characterized by switching behavior in the category purchase decision than by the alternation between inertial and variety seeking states in households' brand choice decisions.[50] Also note that the mean of the inventory parameter is negative and significant in Benchmark 3 and this indicates that Benchmark 3 also suffers from the inventory endogeneity problem.

We now discuss results from the proposed model. The estimate of the price coefficient is not significantly different from that of the other benchmark models. The estimated coefficients for the control functions (we do not show all these estimates for reasons of space) imply negative correlation between prices and the unmeasured

---

[49] We estimated the proposed model without household-specific information $H_h$. This model is also significantly better than Benchmark 3 in all three fit measure. Thus, we conclude that the better performance of the proposed model compared to Benchmark 3 is not due to the household-specific information $H_h$.

[50] The result of Benchmark 3 implies that consumers may alternate between strongly inertial periods characterized by large positive state dependence effect and weakly inertial periods characterized by small positive state dependence effect.

product characteristics, as in the other benchmark models. Households show inertia in their brand choice behavior and this is consistent with the general findings in the literature that low-priced, frequently purchased grocery categories are characterized by inertia (Erdem 1996; Keane 1997; Roy et al. 1996; Seetharaman and Chintagunta 1998).

The parameters related with individual-level dynamics ($\bar{\alpha}_1$ and $\sigma_\alpha$) are significant. These estimates imply that consumers switch between states of high and low purchase tendencies which are characterized by $\bar{\alpha}_0 = 0$ and $\bar{\alpha}_1 = 2.15$, respectively. Specifically, the probability of "No purchase" is 0.60 when $\bar{\alpha}_0 = 0$ and 0.91 when $\bar{\alpha}_1 = 2.15$, on average. Equivalently, the probability of purchasing yogurt on a given visit is 0.40 when a household has a high purchase tendency and is 0.09 when it has a low purchase tendency. Estimates of $p_1$, $p_2$, $q_1$, and $q_2$ tell us that $H_h$, or the number of yogurt purchases during the sample period, is significantly positively related to the probability of being in a high purchase tendency state but not significantly related to the probability of being in a low purchase tendency state. That is, frequent buyer households are more likely to be in a high purchase tendency state, a finding that is intuitive.

By investigating the smoothed probability $\{P(S_{ht} \mid \psi_{hT_h})\}_{t=1}^{T_h}$, we obtain valuable insights into individual-level parameter dynamics, or the households' switching behaviors between high and low purchase tendencies. For the classification of each household's state at $t$, we assume that household $h$ has high purchase tendency if $P(S_{ht} = 1 \mid \psi_{hT_h})$ is significantly smaller than 0.5 and has low purchase tendency if $P(S_{ht} = 1 \mid \psi_{hT_h})$ is significantly larger than 0.5.[51] Based on these rules, 19.9% of the observations (or 17,234 visits) are classified as high purchase tendency state, 66.9%

---

[51] Recall that we can derive the confidence intervals of smoothed probabilities using the bootstrap method described in section 3.3. We use a 90% confidence interval to determine the significance.

(or 57,879 visits) as low purchase tendency state, and the remaining 13.2% (or 11,379 visits) as indeterminate state (i.e. $P(S_{ht} = 1 | \psi_{hT_h})$ is not different from 0.5). Average durations of high and low purchase tendency states are 52.9 visits (or 33 weeks) and 99.7 visits (or 62 weeks), respectively. Among the 519 households, 212 households (40.8%) show switching behaviors between high and low purchase tendencies during the sample period (we call this group "Switching"). Within these households, average durations of high and low purchase tendency states are 30.1 visits (or 19 weeks) and 71.1 visits (or 44 weeks). 221 households (42.5%) have low purchase tendencies, and 78 households (15.0%) have high purchase tendencies, throughout the sample period (we call these groups "Low" and "High" respectively). The proposed model's improvement in model fit compared to the static parameter models (i.e. Benchmark 1 and Benchmark 2) is mainly attributable to the substantial proportion of households that show switching behaviors. Figure 3.2 shows $\{P(S_{ht} = 1 | \psi_{hT_h})\}_{t=1}^{T_h}$ with 90% confidence bands for four representative households. We can observe that the first two households (HHID=3103432 and 3104976) show switching behaviors between high and low purchase tendencies. The third household (HHID=3100008) has a low purchase tendency during the entire sample period. The last household (HHID=3109140) has a high purchase tendency during the entire sample period.

The mean and heterogeneity parameters of inventory (INV) are both significant. In particular, most households have positive coefficients, which imply that the larger the inventory before the shopping trip, the higher the utility of "No purchase." This result is intuitive and in line with previous empirical studies. Note that the results from all three benchmark models indicate the opposite: the estimated effects of inventory are all significantly negative. The proposed model captures the unobservable substitute inventory using regime-switching variables and thereby overcomes the inventory endogeneity problem.
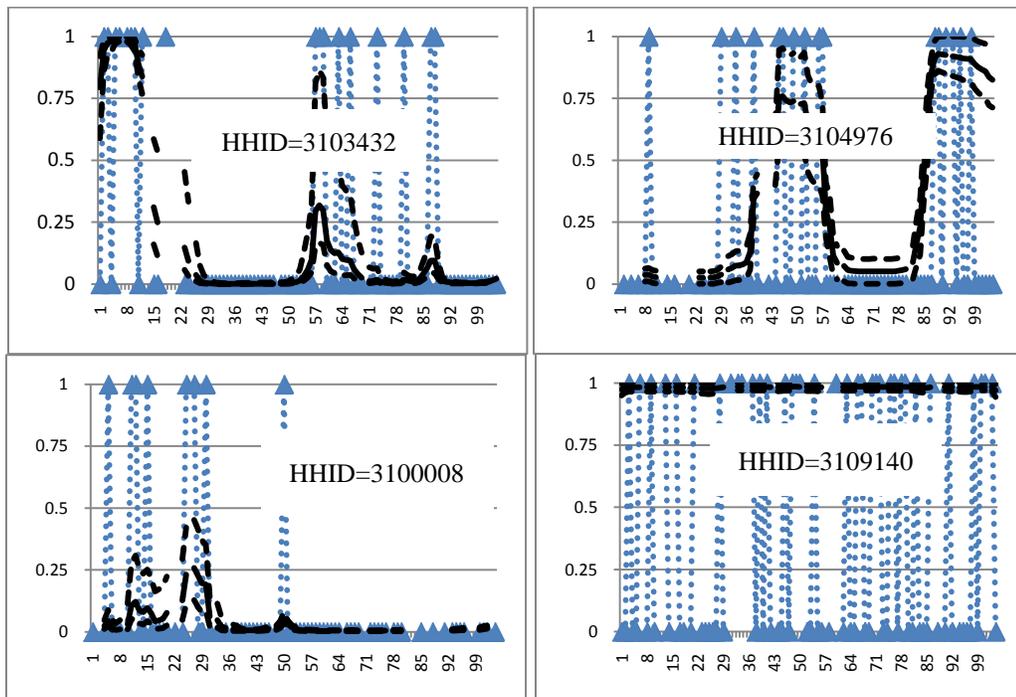
**Figure 3.2: Examples of smoothed probability**

**x-axis: weeks; y-axis: smoothed probability of Purchase in category;**

**Triangles indicate observed Purchases (=1) and No-purchases (=0)**

### 3.4.3 Managerial Application: Optimal Timing of Targeted Promotions

The proposed model captures households' switching behaviors between high and low purchase tendencies by introducing a Markov-switching term in the latent utility. In the empirical application to yogurt data the model is preferred to three benchmark models that are based on extant approaches. What are the managerial implications of such switching behavior of consumers? We answer this question by identifying a unique opportunity for targeted promotions that is implied by the results of our study. In this section, we show via simulation that an understanding of the *dynamic nature* of high and low purchase tendencies is useful in deciding the *optimal customized timing* of a targeted price promotion. While targeting has been the subject of intensive study in the promotions literature, we believe the question of optimal timing has remained hitherto underexplored.

We focus here only on those households that show switching behavior in their yogurt purchase tendencies (40.8% of the sample households), since identification of this group is a contribution of our model. For this illustration, we assume the role of a product manager of Yoplait who is planning a targeted consumer price promotion – a one-time 30% price-off. Our goal is to measure the impact of the temporary price reduction on the choice probability of Yoplait when the *timing* of the offer is *customized to each consumer* based on our knowledge of that household's time-varying category purchase tendencies. This idea is explained in detail next.

Broadly, the question of interest in our two-regime model is whether it is better to offer the price discount when a consumer is in a state of high or low yogurt purchase tendency. We assume the product manager's objective is to accomplish the greatest increase in unit sales of Yoplait, which is equivalent to maximizing the absolute increase in the choice probability of Yoplait (as indicated by the derivative of

the choice probability)[52].   For simplification we do not consider competitive response

to Yoplait's move, although we believe that would be an important extension to

consider in future research.

For Yoplait ($j$=1), the choice probability and derivative of choice probability

with respect to price are as follows:

$$\Pr(y^s_{ht} = 1 \mid S_{ht}) = \frac{\exp(V^s_{h1t})}{\exp(\overline{\alpha}_0 S_{ht} + \overline{\alpha}_1 (1 - S_{ht}) + a_h + INV_{h,t}\beta_{INV,h}) + \sum_{k=1}^{J} \exp(V^s_{hkt})}$$

(30)

$$deriv^s_{h1t} \mid S_{ht} = \frac{\partial \Pr(y^s_{ht} = 1 \mid S_{ht})}{\partial P^s_{1t}} = \Pr(y^s_{ht} = 1 \mid S_{ht})(1 - \Pr(y^s_{ht} = 1 \mid S_{ht}))\beta_{P,h} \quad (31)$$

Note that $\left| (deriv^s_{h1t} \mid S_{ht} = 0) \right| > \left| (deriv^s_{h1t} \mid S_{ht} = 1) \right|$ when $\Pr(y^s_{ht} = 1 \mid S_{ht} = 0) < 0.5$.[53]

Recalling that $S_{ht} = 0$ indicates a regime of high yogurt purchase tendency, this

inequality tells us that the price promotion achieves bigger increases in choice

probability during a high purchase tendency period.  We illustrate this with a

numerical example.  Assume that $\Pr(y^s_{ht} = 1 \mid S_{ht} = 0)$=0.44 and $\Pr(y^s_{ht} = 1 \mid S_{ht} = 1)$

=0.08.  With a price discount of 10%, $\Pr(y^s_{ht} = 1 \mid S_{ht} = 0)$=0.49 and $\Pr(y^s_{ht} = 1 \mid S_{ht} = 1)$

=0.10.  The absolute change in choice probability of Yoplait is larger when $S_{ht} = 0$

(i.e. 0.49 - 0.44 = 0.05) than when $S_{ht} = 1$ (i.e. 0.10 - 0.08 = 0.02), as predicted by the

derivative.  Intuitively, this result implies that price discounts or other promotional

activities are more effective when a consumer is more likely to purchase in the

category, conditional on the brand choice probability being less than 0.5.  This

conclusion is consistent with the commonly observed phenomenon of more

promotions being run in high seasons than in low seasons.

---

[52] We also examined the elasticity of the choice probability with respect to price, but concluded that the derivative is the appropriate criterion.
[53] This is the case in our empirical analysis and typical brand choice models that include a "No purchase" option.

To verify this result in our sample yogurt data and to illustrate the importance of understanding the *dynamic nature* of high and low purchase tendencies, we compare the performance of two price promotion strategies using estimation results of the proposed model. Strategy 1: offer a 30% price reduction once in a week when a household has high purchase tendency; Strategy 2: provide a 30% price reduction once in a randomly chosen week (i.e., this is not a customized strategy). Strategy 2 is implemented with 100 random choices of promotion timing. Along with the two strategies based on the results of the proposed model, we perform a similar counterfactual simulation using the results of Benchmark 1. Recall that Benchmark 1 is a random coefficient logit model with lagged incidence variable and unmeasured product characteristics but there are no individual-level dynamics in the model. Since lagged incidence was found to have a negative effect on the utility of no-purchase, it follows that purchase incidence probability is high immediately following a purchase. Accordingly, we assume that each household is provided with a 30% price reduction once in a randomly chosen week immediately following a category purchase.

Table 3.6 summarizes the results of the three simulation studies. Strategy 1 attains an incremental choice probability for Yoplait of 53.1%, while Strategy 2 attains only 27.6% (on average across the 100 replications). This tells us that a brand manager can increase the impact of a price promotion from 27.6% to 53.1% – a gain of 92% – *merely by optimizing the timing of the price promotion*. In Benchmark 1, the incremental choice probability for Yoplait is 36.4% which is better than Strategy 2 but considerably worse than Strategy 1.

**Table 3.6: Comparison of Price Promotion Strategies**

|  | Proposed Model | | Benchmark 1 | |
| --- | --- | --- | --- | --- |
|  | Strategy 1 | Strategy 2 | Promotion at 100 random times | |
| incremental choice probability | 53.1% | Average 27.6% | Average 36.4% | |
|  |  | SD 8.4% | SD 8.9% | |

### 3.4.4 Model Implementation Issues

From the perspective of managers, the model we have proposed is quite challenging to estimate. In this section we consider an approach to make the model more managerially useful. To begin with, we examine if we can predict the classification of households into Low, High, and Switching groups using descriptive statistics of observed purchase histories as proxies, rather than the proposed regime switching model. For this analysis we divided the original 104-week data into an estimation sample (first 18 months) and hold-out sample (last six months). The proposed model was then estimated on the estimation sample and households were classified into the three groups using the model estimates. In Figure 3.3 we show a scatter-plot of the three groups of households with respect to the mean and standard deviation of their inter-purchase times. This analysis shows that these two variables are likely to be good predictors of households' membership into the three groups. A discriminant model is fit to predict households' membership in the three groups using household-level average inter-purchase times and household-level standard deviation of inter-purchase times as predictors. The estimated discriminant model shows good classification ability; the correct classification rate of the model at 63.5% vastly exceeds that due to chance (proportional chance criterion yields 37.4%).
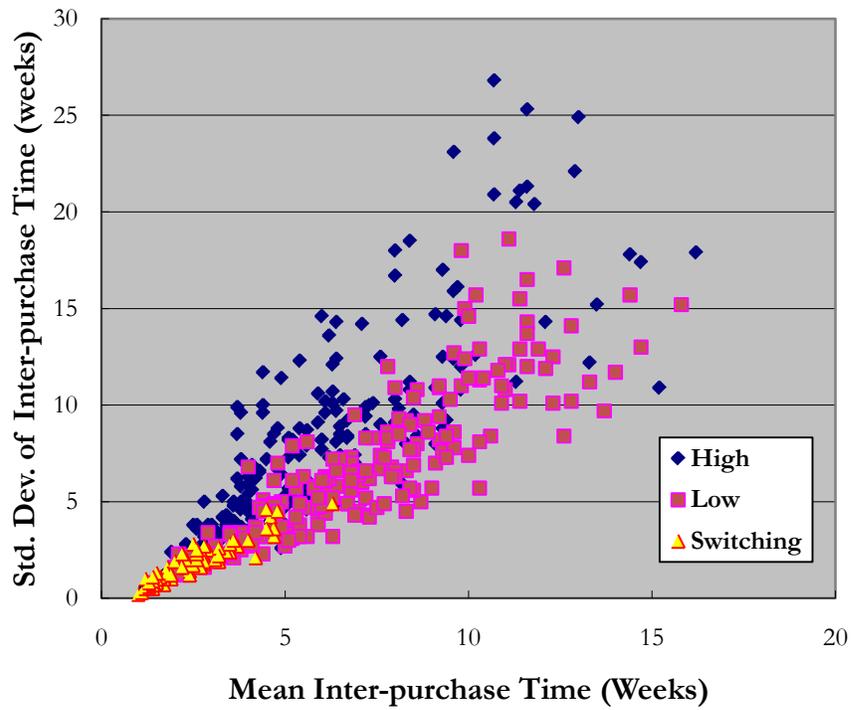
**Figure 3.3: Scatterplot of Mean and Standard Deviation of Inter-Purchase Times by Group**

Next we simulate the effect of a one-off 30% temporary price reduction on Yoplait in the six-month hold-out sample data, similar to the exercise described in the previous section. We target 130 households that have been predicted to belong to the "Switching" group by the discriminant model. Note that because of classification error, only 65% or 80 of these 130 households were identified by the Markov switching model as Switching and the rest were identified as High or Low. Strategy 1 is to offer the price reduction to a household when it is in a high purchase tendency state. To predict the state in the absence of a model, we adopt the following rule-of-thumb: if a household is seen to purchase yogurt on two consecutive store visits, we assume the household is in a high state. Thus, Strategy 1 is to offer a household the price reduction on the visit immediately following two consecutive yogurt purchase visits. As before, Strategy 2 provides a price reduction once in a randomly chosen week (i.e., this is not a customized strategy). Strategy 2 is implemented with 100 random choices of promotion timing. The incremental choice probability under Strategy 1 is found to be 30.5%, while under Strategy 2 is found to be 20% (average across 100 trials). Thus, there is a 50% improvement in performance. Importantly, in this exercise we did not use the estimates from the proposed model, only the conceptual learning that was derived from the model, descriptive statistics of the households' purchasing histories, and a rule-of-thumb to determine when a household is in a high state.

## 3.5 Conclusions and Future Research

We develop a Markov regime-switching random coefficient logit model and apply it to investigate consumers' alternating behavior between high and low category purchase tendencies. We find that scanner panel data in the yogurt category can be better explained by introducing switching levels in the latent utility of "No purchase" than by a static model, after controlling for the influence of marketing mix variables,

inventory, and state dependence.  We also find that 40.8% of sample households switched between high and low category purchase tendencies during the 104-week sampling period.

From a methodological point of view, we propose a unique Markov regime-switching random coefficient logit model which incorporates individual-level parameter dynamics.  In the empirical application of the proposed model, we show that individual-level dynamics are crucial in explaining individuals' idiosyncratic alternation between high and low category purchase tendencies.  Also, we show that alternations between high and low purchase tendencies approximate substitute inventory and if one ignores these, an inventory endogeneity problem occurs and consequently results in biased estimates.  Moreover, to our knowledge the proposed model is the first time-varying parameter discrete choice model that considers unmeasured product characteristics and endogeneity of marketing mix variables.

We demonstrate a managerial application of our proposed model to the targeting of customized price promotions.  Since 40.8% of sample consumers move between states of high and low category purchase tendencies, their response to a targeted price promotion varies depending on when they receive the promotional offer.  We find that offering promotions to consumers when they have a high purchase tendency enhances the effectiveness of the promotion.  In the yogurt data, we show that a brand manager can increase the impact of a price promotion by 92% merely by optimizing the timing of price reduction.  We believe this finding is noteworthy since it introduces a new dimension to targeted marketing decisions – *timing*.  We also show that a firm can implement targeted promotions with customized timing using easily available descriptive statistics of households' purchasing histories.

Several directions exist for further research.  Investigating the household's switching behaviors between high and low purchase tendency over multiple categories

is an important area for additional research. Dynamics in several categories may be correlated and this knowledge will be valuable in understanding and predicting consumers' category purchase decisions. Given the nature of our data, our investigation of a consumer's switching behavior remains at a correlational level. In future research, it will be important to explain a consumer's switching behavior in category purchase at the causal level and identify the underlying mechanisms of such behavior.

Investigating the structural break in the parameters of the random coefficient logit model with an unknown break date is another important area of additional research. Even when researchers do not know the exact date of a consumer's preference change, they usually make the implicit assumption that the change in a consumer's taste coincides with known changes in the marketplace (such as entry of a competitor, or the introduction of a new product). The influence of this assumption on consistency in parameter estimation is an important question in marketing research. Since our model allows the timing of the structural break to vary across individuals and examines the relationship between this behavior and hypothetical explanatory variables, we may investigate the heterogeneity in the structural break of consumer preferences. We can examine whether individuals display a structural break in their preferences, the point of time at which they changed, and which variables correlate with such changes.

REFERENCES

Ailawadi, K. L., S. A. Neslin (1998) The Effect of Promotion on Consumption:
Buying More and Consuming It Faster, Journal of Marketing Research, 35(3),
390-398.

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

Bawa, K. (1990), Modeling Inertia and Variety Seeking Tendencies in Brand Choice
Behavior, Marketing Science, 9(3), 263-278.

Berlyne, D. E. (1970), Novelty, Complexity and Hedonic Value, Perception and
Psychophysics, 8(5A), 279-286.

Berry, S. T. (1994), Estimating Discrete-choice Models of Product Differentiation,
RAND Journal of Economics, 25(2), 242–262.

Bronnenberg, J.B., M.W. Kruger, C.F. Mela (2008), The IRI Marketing Data Set,
Marketing Science, 27(4), 745-748.

Bucklin, R.E., J.M. Lattin (1991), Two-State Model of Purchase Incidence and Brand
Choice, Marketing Science, 19(Winer), 24-39.

Chib, S. (1998), Estimation and Comparison of Multiple Change-point Models,
Journal of Econometrics, 86, 221-241.

Chintagunta, P. K., A. Bonfrer, I. Song (2002), Investigating the Effects of Store-
Brand Introduction on Retailer Demand and Pricing Behavior, Management
Science, 48(10), 1242-1267.

Chintagunta, P. K., D. Jain, N. Vilcassim (1991), Investigating Heterogeneity in Brand
Preferences in Logit Models for Panel Data, Journal of Marketing Research,
28(4), 417–428.

Chintagunta, P. K., J.-P. Dubé (2005), Estimating a Stockkeeping-Unit-Level Brand

151

Choice Model That Combines Household Panel Data and Store Data, Journal of Marketing Research, 42(August), 368-379.

Chintagunta, P. K., J.-P. Dubé, K. Y. Goh (2005), Beyond the Endogeneity Bias: The Effect of Unmeasured Brand Characteristics on Household-Level Brand Choice Model, Management Science, 51(5), 832–849.

Coombs, C. H., G. S. Avrunin (1977), Single Peaked Functions and the Theory of Preference, Psychological Review, 84, 216-230.

Draganska, M., D. C. Jain (2002), A Likelihood Approach to Estimating Market Equilibrium Models, Management Science, 50(5), 605–616.

Erdem, T. (1996), A Dynamic Analysis of Market Structure based on Panel Data, Marketing Science, 15(4), 359-378.

Erdem, T., S. Imai and M. P. Keane (2003), Brand and Quantity Choice Dynamics under Price Uncertainty, Quantitative Marketing and Economics, 1(1), 5-64.

Goolsbee, A., A. Petrin (2004), The Consumer Gains from Direct Broadcast Satellites and the Competition with Cable TV, Econometrica, 72(2), 351–381.

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

Gupta, S. (1988), Impact of Sales Promotions on When, What, and How Much to Buy, Journal of Marketing Research, 25(November), 342-355.

Hamilton, J. D. (1989), A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle, Econometrica, 57(2), 357-384.

Heckman, J. J. (1981), Statistical Models for Discrete Panel Data, C. F. Manski, D. McFadden, eds. Structural Analysis of Discrete Data with Applications, MIT Press, Cambridge, MA, 179–195.

Jeuland, A. P. (1979), Brand Choice Inertia as one Aspect of the Notion of Brand Loyalty, Management Science, 25(4), 671-682.

Jones, J., T. Landwehr (1988), Removing Heterogeneity Bias from Logit Model Estimation, Marketing Science, 7(1), 41-59.

Keane, M. P. (1993), Simulation Estimation for Panel Data Models with Limited Dependent Variables, in The Handbook of Statistics & Econometrics, eds. C. R. Rao, G. S. Maddala, D. Vinod, eds., Elsevier Science Publishers, Amsterdam, The Netherlands, 545–572.

Keane, M. P. (1997), Modeling Heterogeneity and State Dependence in Consumer Choice Behavior, Journal of Business and Economic Statistics, 15(3), 310-327.

Kim, C.-J. (1994), Dynamic Linear Models with Markov-switching, Journal of Econometrics, 60, 1-22.

Kim, J. G., U. Menzefricke, F. M. Feinberg (2005), Modeling Parametric Evolution in a Random Utility Framework, Journal of Business and Economic Statistics, 23(3), 282-294.

Krishnamurthi, L., S.P. Raj (1991), An Empirical Analysis of the Relationship between Brand Loyalty and Consumer Price Elasticity, Marketing Science, 10 (2), 172-183.

Kuksov, D., J. M. Villas-Boas (2008), Endogeneity and Individual Consumer Choice, Journal of Marketing Research, 45 (6), 702-14.

Lachaab, M., A. Ansari, K. Jedidi, A. Trabelsi (2006), Modeling Preference Evolution in Discrete Choice Models: A Bayesian State-space Approach, Quantitative Marketing and Economics, 4, 57-81.

McAlister, L. (1982), A Dynamic Attribute Satiation Model for Choices made Across Time, Journal of Consumer Research, 9(3), 141-150.

Montgomery, A. L., S. Li, K. Srinivasan, J. C. Liechty (2004), Modeling Online Browsing and Path Analysis Using Clickstream Data, Marketing Science, 23(4), 579–595.

Neal, D.T., W. Wood, J. M. Quinn (2006), Habits – A Repeat Performance, Current Directions in Psychological Science, 15, 4, 198-202.

Netzer, O., J. M. Lattin, V. Srinivasan (2008), A Hidden Markov Model of Customer Relationship Dynamics, Marketing Science, 27, 185-204.

Park, S., S. Gupta (2008), A Simulated Maximum Likelihood Estimator for the Random Coefficient Logit Model Using Aggregate Data, Journal of Marketing Research, forthcoming.

Petrin, A., K. Train (2004), Omitted Product Attributes in Discrete Choice Models, working paper, University of Chicago, Chicago, IL.

Ratner, R. K., B. E. Kahn, D. Kahneman (1999), Choosing Less-Preferred Experiences for the Sake of Variety, Journal of Consumer Research, 26(1), 1-15.

Revelt, D., K. Train (2000), Specific Taste Parameters and Mixed Logit, Working Paper, Department of Economics, University of California, Berkeley, CA.

Roger, R. D. (1979), Commentary on 'The Neglected Variety Drive', Journal of Consumer Research, 6(June), 88-91.

Roy, R., P. K. Chintagunta, S. Haldar (1996), A Framework for Investigating Habits, 'the hand of the past,' and Heterogeneity in Dynamic Brand Choice, Marketing Science, 15(3) 280–299.

Seetharaman, P. B. (2003), Probabilistic Versus Random Utility Models of State Dependence: An Empirical Comparison, International Journal of Research in Marketing, 20(1), 87–96.

Seetharaman, P. B., A. K. Ainslie, P. K. Chintagunta (1999), Investigating Household State Dependence Effects Across Categories, Journal of Marketing Research, 36(4), 488-500.

Seetharaman, P. B., P. K. Chintagunta (1998), A Model of Inertia and Variety-seeking with Marketing Variables, International Journal of Research in Marketing,

15(1), 1–17.

Shugan, S. (1980), The Cost of Thinking, Journal of Consumer Research, 7, 99-111.

Singh, V. P., K. T. Hansen, R. C. Blattberg (2006), Market Entry and Consumer Behavior: An Investigation of a Wal-Mart Supercenter, Marketing Science, 25(5) 457–476.

Sudhir, K. (2001), Competitive Pricing Behavior in the US Auto Market: A Structural Analysis, Marketing Science, 20(1), 42-60.

Villas-Boas, J. M., R. Winer (1999), Endogeneity in Brand Choice Models, Management Science, 45(10), 1324-1338.

Villas-Boas, J. M., Y. Zhao (2005), Retailer, Manufacturers, and Individual Consumers: Modeling the Supply Side in the Ketchup Marketplace, Journal of Marketing Research, 42(1), 83-95.

Yang, S., Y. Chen, G. M. Allenby (2003), Bayesian Analysis of Simultaneous Demand and Supply, Quantitative Marketing and Economics, 1(3), 244–264.

APPENDIX 1

STATISTICAL PROPERTIES OF THE PROPOSED SIMULATED MAXIMUM

LIKELIHOOD ESTIMATOR

## A1.1 Introduction

In this Appendix we demonstrate some statistical properties of the simulated maximum likelihood estimator proposed in Chapter 1.1. As described in Chapter 1.1, our estimation approach involves drawing a new sample of size $R$ (smaller than $H$) based on shares in the observed sample of size $H$, in order to overcome numerical difficulties associated with using the observed sample, if $H$ is large. Thus, our sample of size $R$ could be considered a "two-stage sample". The alternative is a sample of size $R$ taken directly from the population (call this a "one-stage sample"). We know that standard results, such as consistency, apply to SML estimates obtained from a one-stage sample. We determine some of the properties of the two-stage sample by comparing a one-stage sample with a two-stage sample, both of size $R$. If the two samples are very similar, we expect standard results to apply to the two-stage sample as well.

In this Appendix we proceed as follows. First we define a two-stage sample. Next, we use simulations to examine the similarity between one-stage and two-stage samples. Finally, we directly compare SML estimates obtained from one-stage and two-stage samples for different values of $R$ in simulated data.

## A1.2 Definition of two-stage sample

Define

$$P_{jt} = \int \frac{\exp(x_{jt}\beta_h + \xi_{jt})}{1 + \sum \exp(x_{it}\beta_h + \xi_{it})} \phi(\beta)\, d\beta \, .$$

Let us assume that we observe aggregate shares $\{S_{jt}\}(= N_{jt}/H)$ based on a sample of size $H$ which is finite, and thus observed shares contain sampling error. $\{N_{jt}\}$ ($\Sigma_j N_{jt} = H$) are outcomes of $H$ multinomial draws from probabilities $\{P_{jt}\}$. When $\{N_{jt}\}$ are large, numerical problems prevent direct application of SML. In this case, we perform $R$ multinomial draws ($R<H$) from shares $\{S_{jt}\}$ and get outcomes $\{M_{jt}\}$ ($\Sigma_j M_{jt} = R$, define $Q_{jt} = M_{jt}/R$). This is our two-stage sample. By using $\{M_{jt}\}$ instead of $\{N_{jt}\}$ to obtain SML estimates, we can circumvent the numerical difficulty of handling large exponents.

**A1.3 Comparison of two-stage sample with one-stage sample**

We use the DGP of case 1 of the simulation study reported in Chapter 1.1. We calculate $\{P_{jt}\}$ using 100,000 random draws and set $H$=1,000 and $R$=100. To generate a one-stage sample we perform $R$(=100) multinomial draws from $\{P_{jt}\}$. To generate a two-stage sample we perform $H$(=1,000) multinomial draws from $\{P_{jt}\}$, calculate $\{S_{jt}\}(= N_{jt}/H)$, and then perform $R$(=100) multinomial draws from $\{S_{jt}\}$ to obtain $\{M_{jt}\}$ or $\{Q_{jt} = M_{jt}/R\}$. For each set of $\{P_{jt}\}$, we generated 100 sets of one-stage samples and 100 sets of two-stage samples. So, for $P_{jt}$ with each $j$ and $t$, we have $Q_{jt,d}^{1-stage}$ ($d$=1,…,100) from one-stage sample and $Q_{jt,d}^{2-stage}$ ($d$=1,…,100) from two-stage sample.
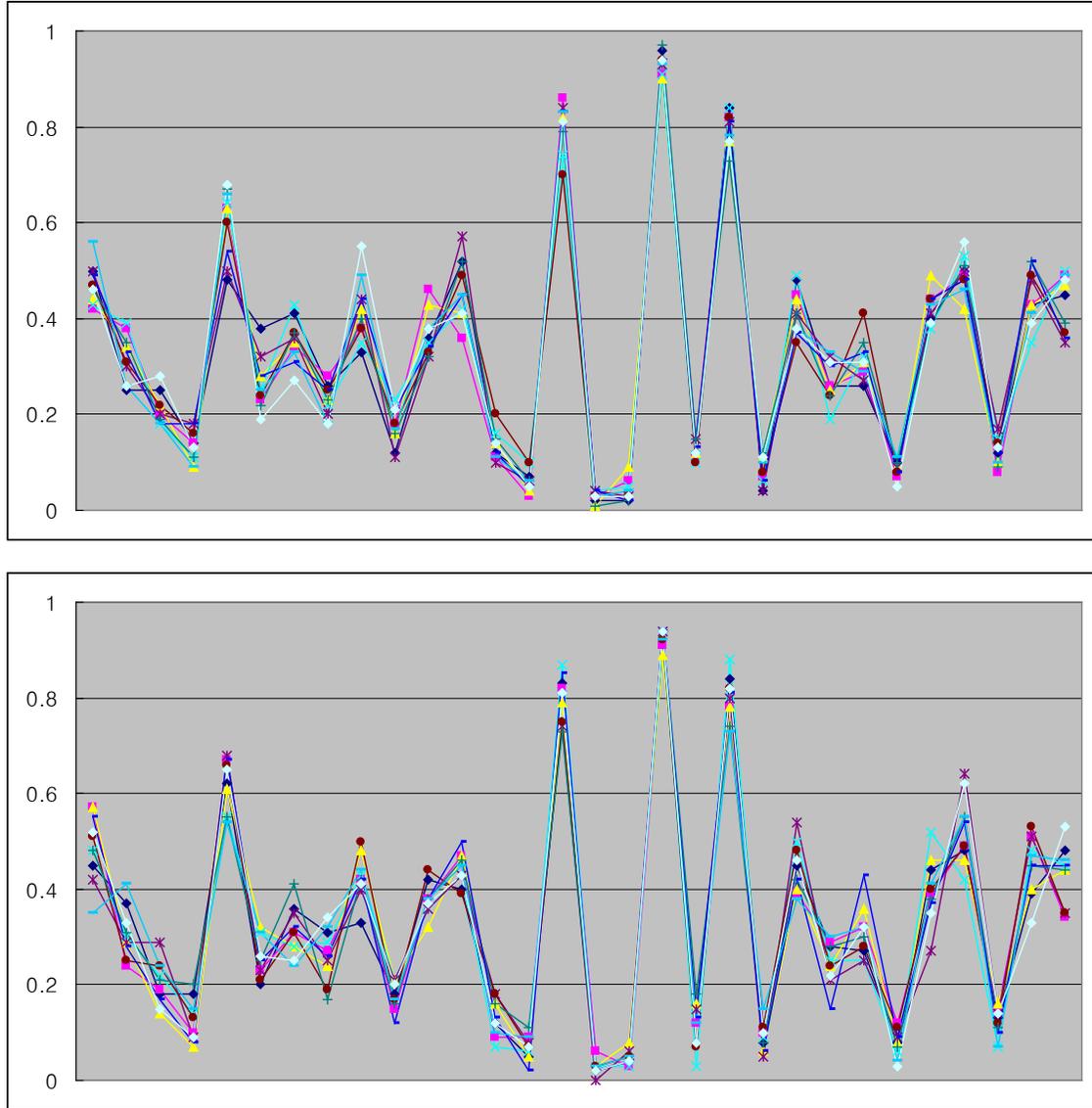
**Figure A1.1: One-stage samples (top panel) vs. Two-stage samples (bottom panel)**

**x-axis: ($t$=1; $j$=1), ($t$=1; $j$=2), ($t$=1; $j$=3), ($t$=2; $j$=1), ($t$=2; $j$=2), ($t$=2; $j$=3), y-axis:**

$Q_{jt,d}^{1-stage}$ **(top panel) and** $Q_{jt,d}^{2-stage}$ **(bottom panel) for $d$=1,…,10.**

In Figure A1.1 we show $Q_{jt,d}^{1-stage}$ and $Q_{jt,d}^{2-stage}$ for the first ten samples, ($d=1,\ldots,10$). (We do not show all 100 samples to allow easier reading of the figure.) We find that both $Q_{jt,d}^{1-stage}$ and $Q_{jt,d}^{2-stage}$ are centered around $\{P_{jt}\}$ and their distributions look quite similar.

Next we calculate the average absolute distances of $Q_{jt,d}^{1-stage}$ and $Q_{jt,d}^{2-stage}$ respectively from $P_{jt}$. We define

$$ dist\_1-stage = \frac{\Sigma_{t=1\sim T}\Sigma_{j=1\sim J}(\Sigma_{d=1\sim100}\,|\,P_{jt}-Q_{jt,d}^{1-stage}\,|\,/100)}{J*T} $$

$$ dist\_2-stage = \frac{\Sigma_{t=1\sim T}\Sigma_{j=1\sim J}(\Sigma_{d=1\sim100}\,|\,P_{jt}-Q_{jt,d}^{2-stage}\,|\,/100)}{J*T} $$

In our data $dist\_1-stage$ =0.033 and $dist\_2-stage$ =0.034. Table A1.1 summarizes the results from simulations with different values of $H$ and $R$. In general, $dist\_1-stage$ is very close to $dist\_2-stage$. As $R$ increases, both distances decrease. Comparing results for $H$=10,000 and $H$=1,000, we observe that $dist\_2-stage$ decreases only marginally as $H$ increases. From all these results, we conclude that a one-stage sample is empirically equivalent to a two-stage sample and thus we expect that the standard results of SML apply to a two-stage sample.

**Table A1.1:  Results of Simulation Study**

| H | R | dist_2-stage | dist_1-stage |
|---|---|---|---|
| 10,000 | 150 | 0.027 | 0.027 |
| 10,000 | 100 | 0.033 | 0.033 |
| 10,000 | 50 | 0.046 | 0.046 |
| 1,000 | 150 | 0.028 | 0.027 |
| 1,000 | 100 | 0.034 | 0.033 |
| 1,000 | 50 | 0.047 | 0.046 |

**Table A1.2: Results of Simulation Study to Compare SML estimates from One-stage versus Two-stage Samples**

| | | Two-stage Sample | | | | | |
|---|---|---|---|---|---|---|---|
| Parameters | True values | H=1000, R=50 | | H=1000, R=100 | | H=1000, R=150 | |
| | | Mean | SE | Mean | SE | Mean | SE |
| $\bar{\beta}_1$ | 0.2 | 0.211 | 0.100 | 0.190 | 0.100 | 0.185 | 0.088 |
| $\bar{\beta}_2$ | 0.5 | 0.515 | 0.102 | 0.512 | 0.095 | 0.492 | 0.090 |
| $\bar{\beta}_3$ | -1 | -1.013 | 0.114 | -0.976 | 0.105 | -0.985 | 0.090 |
| $\bar{\beta}_4$ | 1 | 0.988 | 0.144 | 0.980 | 0.122 | 0.991 | 0.093 |
| $\theta_1 = \theta_2$ | 1 | 1.006 | 0.101 | 0.996 | 0.111 | 1.028 | 0.111 |
| $\sigma_{33}$ | 1 | 1.011 | 0.133 | 1.000 | 0.122 | 1.005 | 0.109 |
| $SD(\omega_{2jt})$ | 0.707 | 0.710 | 0.069 | 0.736 | 0.063 | 0.755 | 0.058 |

| | | One-stage Sample | | | | | |
|---|---|---|---|---|---|---|---|
| Parameters | True values | R=50 | | R=100 | | R=150 | |
| | | Mean | SE | Mean | SE | Mean | SE |
| $\bar{\beta}_1$ | 0.2 | 0.206 | 0.095 | 0.191 | 0.092 | 0.196 | 0.089 |
| $\bar{\beta}_2$ | 0.5 | 0.509 | 0.100 | 0.507 | 0.090 | 0.503 | 0.088 |
| $\bar{\beta}_3$ | -1 | -0.984 | 0.110 | -0.976 | 0.103 | -0.980 | 0.101 |
| $\bar{\beta}_4$ | 1 | 0.979 | 0.138 | 0.980 | 0.138 | 0.984 | 0.132 |
| $\theta_1 = \theta_2$ | 1 | 0.994 | 0.106 | 0.990 | 0.104 | 0.994 | 0.102 |
| $\sigma_{33}$ | 1 | 0.984 | 0.145 | 0.986 | 0.131 | 0.991 | 0.126 |
| $SD(\omega_{2jt})$ | 0.707 | 0.687 | 0.063 | 0.699 | 0.059 | 0.714 | 0.058 |

**A1.4 Statistical properties of the proposed SML estimator**

We consider how SML estimates behave as *R* changes in a simulation study. We use the same DGP as in case 1 of the paper. We generate $\{N_{jt}\}$ with *H*=1,000 and then draw three sets of two-stage samples with *R*=(50, 100, 150) as well as one-stage samples of the same three sizes. We generated 200 data sets for this simulation study.

We expect that standard errors decrease as *R* increases. The rate of decrease is expected to be $\sqrt{N}$ if there is no numerical integration for heterogeneity and unmeasured product characteristics. In the presence of numerical integration, the rate is expected to be slower than $\sqrt{N}$.

The top panel in Table A1.2 summarizes the first two moments of estimates from two-stage samples. All estimates are tightly distributed around the true values. When we compare (*H*=1000, R=50) to (*H*=1000, *R*=100) and (*H*=1000, *R*=150), we observe that for each parameter (with the exception of □ ) SE's decrease as *R* increases. The rate of decrease is slower than $\sqrt{N}$. The lower panel reports the first two moments of estimates from one-stage samples. Here we confirm all the same results as two-stage samples. By comparing upper and lower panels, we find that in all cases (*R*=50, *R*=100, and *R*=150), we obtain comparable results from one-stage and two-stage samples.

# SIMULATION STUDY

We perform a simulation study to show that the proposed model is identified and the suggested estimation method is valid in the recovery of model parameters. The specific Data Generating Process (DGP) and the parameter values we assign are summarized below:

$$U_{h0t} = \alpha_{h,t} + \varepsilon_{h0t}, \quad \text{if no purchase,} \tag{A2.1}$$

$$U_{hjt} = \beta_{j,h} + P_{jt}\beta_{P,h} + \varepsilon_{hjt}, \quad j=1, 2. \tag{A2.2}$$

$$
\begin{bmatrix} \alpha_{h,t} \\ \beta_{1,h} \\ \beta_{2,h} \\ \beta_{P,h} \end{bmatrix} = \begin{bmatrix} \overline{\alpha}_{h,t} \\ \overline{\beta}_1 \\ \overline{\beta}_2 \\ \overline{\beta}_P \end{bmatrix} + \begin{bmatrix} a_h \\ b_{1,h} \\ b_{2,h} \\ b_{P,h} \end{bmatrix} = \begin{bmatrix} \overline{\alpha}_0(1-S_{ht}) + \overline{\alpha}_1 S_{ht} \\ \overline{\beta}_1 \\ \overline{\beta}_2 \\ \overline{\beta}_P \end{bmatrix} + \begin{bmatrix} a_h \\ b_{1,h} \\ b_{2,h} \\ b_{P,h} \end{bmatrix}
$$

$$
= \begin{bmatrix} 0(1-S_{ht}) + 2S_{ht} \\ 0 \\ -0.5 \\ -1 \end{bmatrix} + \begin{bmatrix} a_h \\ b_{1,h} \\ b_{2,h} \\ b_{P,h} \end{bmatrix} \tag{A2.3}
$$

$$
\begin{bmatrix} a_h \\ b_{1,h} \\ b_{2,h} \\ b_{P,h} \end{bmatrix} \sim N(0_4, (0.3)^2 \cdot I_4), \tag{A2.4}
$$

$$P_{jt} \sim N(1.5, (0.3)^2), \tag{A2.5}$$

$$\varepsilon_{h0t} \text{ and } \varepsilon_{hjt} \sim \qquad \text{Type-1 Extreme Value,} \tag{A2.6}$$

$$P(S_{ht} = 0 \mid S_{ht-1} = 0) = q = 0.98, \tag{A2.7}$$

$$P(S_{ht} = 1 \mid S_{ht-1} = 1) = p = 0.98. \tag{A2.8}$$

**Table A2.1: Results of the Simulation Study**

**(based on 100 replications)**

| Parameters | True Value | Mean | SE |
|:---:|:---:|:---:|:---:|
| $\overline{\beta}_1$ | 0.00 | -0.036 | 0.143 |
| $\overline{\beta}_2$ | -0.50 | -0.530 | 0.152 |
| $\overline{\beta}_P$ | -1.00 | -0.945 | 0.105 |
| $\overline{\alpha}_1$ | 2.00 | 2.051 | 0.145 |
| $SD(b_{1,b})$ | 0.30 | 0.274 | 0.097 |
| $SD(b_{2,b})$ | 0.30 | 0.253 | 0.112 |
| $SD(b_{P,b})$ | 0.30 | 0.238 | 0.073 |
| $SD(a_b)$ | 0.30 | 0.195 | 0.126 |
| $q$ | 0.98 | 0.975 | 0.004 |
| $p$ | 0.98 | 0.976 | 0.005 |

The DGP incorporates the individual-level parameter dynamics through $\alpha_{b,t}$ which is governed by a Markov process described in (A2.7) – (A2.8). The heterogeneity of individuals is also considered through (A2.3). In the paper, we proposed a method to handle the unmeasured product characteristics and price endogeneity using control functions and fixed-effects for the exogenous unmeasured product characteristics. In the light of model identification, these are nothing but the additional regressors. Therefore, we do not include them in this simulation study for the simplicity. We set $t = 1,\ldots,100$ for all $h$, and $h=1,\ldots,100$. We generated 100 data sets from the DGP. The first two moments of the empirical sampling distributions of the parameter estimates are summarized in Table A2.1. The proposed method works well in the recovery of all parameters. Even when the sample is as small as $H=100$ and $T=100$, the estimates of parameters are distributed close to the true values and we can conclude that the method provides unbiased estimates.