RESEARCH NOTES AND COMMUNICATIONS

TECK-HUA HO and JUIN-KUAN CHONG*

The authors develop a model to describe and predict consumer stockkeeping-unit choice in frequently bought product categories. The model posits that a product category consists of several salient attributes with many attribute levels and represents a stockkeeping unit as an attribute-level combination. The number of parameters of the model does not increase with the number of stockkeeping units and the number of attribute levels. The authors demonstrate the descriptive and predictive power of their model using 133,492 purchase incidences in 16 product categories. Their model fits 7% better in sample and predicts 8% better out of sample in hit probability than two leading models and requires only half the number of parametes.

A Parsimonious Model of Stockkeeping-Unit Choice

Most consumer product categories have hundreds of stockkeeping units (SKUs), and the number continues to grow (Quelch and Kenny 1994). It is a challenge to estimate most prevailing consumer choice models (e.g., Allenby and Rossi 1991; Erdem and Keane 1996; Guadagni and Little 1983) because the models comprise product-specific parameters that are at least as large as the number of items in the categories.

*Teck-Hua Ho is William Halford Jr. Family Professor of Marketing, Haas School of Business, University of California, Berkeley (e-mail: hoteck@hass.berkeley.edu). Juin-Kuan Chong is Assistant Professor of Marketing, National University of Singapore Business School, National University of Singapore (e-mail: bizcjk@nus.edu.sg). This research is partially supported by National Science Foundation Grant 0078853 and NUS Research Grant R-314-000-018-112. The authors thank Andrew Ainslie, Greg Allenby, David Bell, Eric Bradlow, Colin Camerer, Tulin Erdem, Peter Fader, John Little, Len Lodish, Carl Mela, Alan Montgomery, Gary Russell, David Schmittlein, Richard Staelin, Christophe Van den Bulte, and Russ Winer, as well as seminar participants at Marketing Science Conference at Syracuse University; Duke University; Massachusetts Institute of Technology; National University of Singapore; Singapore Management University; University of Texas, Dallas; Washington University, St. Louis; and the Wharton School, for their helpful suggestions. The authors are especially grateful to John Lynch for his help in developing a behavioral underpinning for their model. David Bell, Pete Fader, and Bruce Hardie generously provided the data. The authors also thank the three anonymous *JMR* reviewers for their many suggestions.

Three approaches have been adopted to overcome this challenge. The first approach reduces the number of product-specific parameters by focusing on a subset of the SKUs. In this approach, either all purchase incidences of the least frequently bought SKUs are discarded (e.g., Fader, Lattin, and Little 1992; Siddarth, Bucklin, and Morrison 1995) or all purchase incidences of consumers who bought the products are discarded (e.g., Chintagunta 1993). Both ways of discarding data amount to choice-based sampling (see Ben-Akiva and Lerman 1985), which can lead to potential bias in product-specific parameters (Manski and Lerman 1977).

The second approach aggregates the level of analysis to a higher level (e.g., from SKU to brand-size combination, as Bucklin and Gupta [1992] and Guadagni and Little [1983] do) or aggregates a subset of products into a composite product (e.g., aggregate several least frequently bought products into a composite "other" product, as Chiang [1991], Erdem and Keane [1996], and Papatla and Krishnamurthi [1992] do). Choice aggregation may lead to biased product-specific parameters (Ben-Akiva and Lerman 1985) if the composition of the member products changes over time as a result of varying product availability.[1] Product availability can vary widely because of stock-out and peri-

[1]If the existence of a nested logic structure in the consumer choice process can be assumed, aggregation of the level of analysis may not lead to biased estimates. We thank an anonymous reviewer for bringing this to our attention.

Journal of Marketing Research
Vol. XL (August 2003), 351–365

odic product-line extensions and deletions. For example, during a two-year period, our data set had a minimum of 8 extensions and 8 deletions in the egg category and as many as 141 extensions and 127 deletions in the detergent category.

Besides producing potentially biased estimates, the previous two approaches do not estimate demand for all SKUs, which is essential for effective inventory planning and shelf-space allocation. For this reason, operations management researchers have found it difficult to incorporate scanner-based choice models into their inventory planning tools (Ho and Tang 1998). Our model aims to address this deficiency. As Chong, Ho, and Tang (2001) show, such a microlevel model can be useful in selecting optimal product assortment.

The third approach overcomes these limitations by assuming that a product category comprises a small number of salient attributes and that each salient attribute has different levels. An SKU derives its intrinsic value from the attribute levels it assumes (Fader and Hardie 1996). The product-specific parameters become sums of attribute-level values. This approach works if the total number of attribute levels for each salient attribute is small; however, this is often not the case. Of the product categories in our data set, 9 of 16 have more than 90 attribute levels (see Table 1).

In this article, we develop a model that will work well for large categories with many attribute levels. Building on the work of Guadagni and Little (1983) and Fader and Hardie (1996), we modify the standard utility function commonly used in the scanner-data literature by capturing several behavioral regularities reported in the consumer research. Thus, our approach is behavioral in nature, and our goal is to develop a more predictive utility function. Our approach leads to a parsimonious model, because we specify a dynamic structure to capture the ways a consumer experiences a product and attribute levels over time. In addition, we allow for autoregressive error structures at both the attribute and the product levels.

Our model suggests that a consumer's utility for an attribute level changes over time, because the consumer accumulates a consumption experience for the chosen attribute level and a shopping experience for all familiar and available attribute levels, both of which depend on the associated attribute-level familiarity. The notion of shopping experience is new and, as we show subsequently, is crucial in describing product-choice behavior. The notion also enables us to predict how variety seeking occurs. If shopping experience increases with attribute-level familiarity, familiar attribute levels are more likely than unfamiliar attribute levels to be chosen in the future. Because of stronger shopping experience, a familiar, unchosen attribute level can receive a higher overall experience than a chosen, unfamiliar attribute level. Other things being equal, our model predicts that the consumer is more likely to switch back to products with more familiar attribute levels.

In addition to the attribute-level experience, the consumer develops an idiosyncratic product-specific experience. Similar to attribute-level experience, this product-level experience includes shopping and consumption, and it increases with familiarity. Because product familiarity evolves over time, the consumer responds differently not only to different products on the same purchase occasion but also to the same product over time.

Our model makes three contributions. First, it uses all purchase data, does not aggregate SKUs, and has a parameter number that does not increase with the number of SKUs or attribute levels. For example, if a product has three salient attributes (e.g., brand, size, and flavor), our model has only 59 parameters. In general, our two-segment model has $11 + 12 \times (K + 1)$ parameters for a K-attribute product category. Consequently, it can be used to model product choice in any frequently bought product categories, including those with hundreds of SKUs. More important, our model fits and predicts SKU choice better than the Fader and Hardie (1996; FH) model does. Second, using panel-level data from seven small product categories, we benchmarked our model against the FH and Guadagni and Little (1983; GL) models. On average, our model fits 7% better than does the FH model in-sample and predicts 8% better out-of-sample in hit probability. In terms of adjusted R^2, the model is 8% and 11% higher in-sample and out-of-sample, respectively. The improvement in fit over the GL model is even better. For example, the improvement in adjusted R^2 is 15% in-sample and 19% out-of-sample. In addition, this superior performance is achieved with only half the number of parameters. Third, the model incorporates several behavioral regularities that have been documented in consumer research. We use scanner data to test memory-based grocery shopping (e.g., Alba, Hutchinson, and Lynch 1991; Lynch, Marmorstein, and Weigold 1988) in the field and find strong support for the phenomenon.

In the next section, we present our model. In the section "Empirical Analysis," we test the behavioral premises that underlie our model, provide empirical evidence to substantiate its superior performance (in fit and prediction), and discuss several empirical regularities. We conclude by discussing managerial implications and applications, and we suggest further research directions.

THE SKU CHOICE MODEL

Consider consumer i, who visits a store to buy an SKU in a particular product category. The product category has many SKUs indexed by j. The consumer evaluates the product category by a set of K salient attributes indexed by k. Each salient attribute k has L_k attribute levels indexed by l. Each SKU j offers an attribute-level combination. For example, the ice cream product category may be evaluated by such salient attributes as brand, size, and flavor. The possible attribute levels for ice cream flavor are vanilla, chocolate, strawberry, and so on. An SKU offers an attribute-level combination such as "Ben & Jerry's, 16 oz., vanilla." The consumer makes product purchase on multiple occasions. On each purchase occasion, the consumer decides which SKU j to buy and consume, given all SKUs' marketing-mix activities.[2] The goal of the model is to predict which SKU j the consumer i will choose on a purchase occasion given the purchase history.

[2]It is possible that the consumer purchases multiple SKUs on a particular shopping occasion. In model calibration, the purchases can be treated as either simultaneous or sequential observations in model updating (such as Equations 3 and 4). We adopt the simultaneous approach because it does not use purchase information of a product to predict the purchase of another product on the same store visit. We also estimated the model using the sequential approach and found that both approaches generated similar parameter estimates but (obviously) better fit.

Table 1
DATA DESCRIPTION OF PRODUCT CATEGORIES

	Small Product Categories							*Large Product Categories*								
	Eggs	*Fabric Softener*	*Bathroom Tissue*	*Cola*	*Bacon*	*Paper Towels*	*Hot Dogs*	*Potato Chips*	*Yogurt*	*Spaghetti Sauce*	*Soap*	*Tooth-paste*	*Detergent*	*Regular Cereal*	*Ice Cream*	*Frozen Pizza*
Category Summary																
Total sample size	9903	9781	14,590	14,705	3698	12,218	4111	7022	12,594	4226	5214	2993	7171	12,978	6977	5311
Number of households	482	594	528	429	314	495	334	382	356	320	384	306	471	480	420	337
Number of SKUs	38	59	106	141	62	108	128	285	288	194	243	259	321	242	421	337
Total number of levels in all salient attributes	20	22	40	41	45	53	64	95	96	102	107	119	124	153	191	256
Number of Parameters																
Our model	59	73	59	59	59	59	59	59	59	59	59	59	59	59	59	59
Fader and Hardie (1996)	75	79	115	117	125	141	163	225	227	239	249	273	283	341	417	547
Guadagni and Little (1983)	91	133	227	297	139	231	271	585	591	403	501	533	657	499	857	689
Salient Attribute Description																
Brand																
Total number	12	10	21	17	26	27	38	29	15	41	47	24	41	35	37	40
Example	Crystal Fm. Prv. Label W.R. Valley	Downy Snuggle Bounce	Scottissue Northern Charmin	Coca Cola Pepsi Royal Crown	Oscar Mayer W.Corn King Lazy Maple	Bounty Scottowels Versatile	Oscar Mayer Hygrade W.Corn King	Jays Lays Ruffles	Dannon Yoplait Kemps	Ragu Prego Hunts	Dial Dove Ivory	Crest Colgate Arm & Hammer	Tide Wisk All	Kellogg General Mills Post	Value Pak Häagen-Dazs Dreyers	Tombstone Bravissimo Jacks
Package Size																
Total number	3	4	11	16	7	9	11	34	7	30	42	44	62	73	9	145
Example	12 count 18 count 6 count	Small Medium Large	4 rolls 1 roll 12 rolls	67.6 oz. 288 oz. 144 oz.	16 oz. 12 oz. 24 oz.	1 roll 3 rolls 2 rolls	16 oz. 12 oz. 40 oz.	6.5 oz. 7 oz. 6 oz.	6 oz. 8 oz. 32 oz.	30 oz. 26 oz. 14 oz.	15 oz. 14 oz. 9.5 oz.	6.4 oz. 4.6 oz. 6 oz.	64 oz. 128 oz. 42 oz.	12 oz. 18 oz. 15 oz.	64 oz. 16 oz. 32 oz.	22 oz. 20 oz. 17 oz.
Flavor/Ingredient																
Total number	5	4	8	8	12	17	15	32	74	31	18	51	21	45	145	71
Example	Large Extra large Jumbo	Regular Staingard Light	Unscented Regular Soft scented	Regular Diet Caffeine free	Regular Smoked Hickory smoked	White paper Print Assorted colors	Beef Chicken & pork Pork & turkey	Regular BBQ Sour cream & onion	Plain Strawberry Raspberry	Plain Italian garden Tomato & herb	Regular Original Unscented	Tartar control Baking soda Regular	Regular liquid Concentrated powder Regular powder	Corn Wheat bran Rice	Vanilla Neapolitan Chocolate	Sausage Cheese Deluxe
Formula																
Total number		4														
Example		Regular Staingard Light														

Notes: The product categories are arranged in increasing order of the total number of levels in all salient attributes.

Each purchase occasion defines a time epoch. Each time epoch t begins with store visit, followed by shopping and purchase, and then attribute- and product-level consumption. While shopping, the consumer gathers information on the marketing-mix activities of products on the choice menu. Consumer i's utility for SKU j at the end of shopping (but before purchase) in time epoch t (denoted by $U_{ij}[t]$) is a sum of two components, namely, consumer i's intrinsic value for SKU j (denoted by $V_{ij}[t]$) and the value associated with SKU j's marketing-mix activities (denoted by $M_j[t]$). We have

$$(1)\qquad U_{ij}(t) = V_{ij}(t) + M_j(t) + \varepsilon_{ij}(t),$$

where $\varepsilon_{ij}(t)$ is an aggregate error term that consists of multiple error components, each of which may exhibit a serial correlation across time. We discuss the exact composition of the error term in greater detail subsequently.[3] Most prior models have used the additive form, and we adopt it here for simplicity.

Intrinsic Value of an SKU

The intrinsic value of SKU j is the additive sum of the cumulative attractions of the attribute levels SKU j assumes ($A_{ikl}[t]$) and the product as a whole ($A_{ij}[t]$). Formally,

$$(2)\qquad V_{ij}(t) = \sum_{k=1}^{K}\sum_{l=1}^{L_k} A_{ikl}(t) \times I_{jkl} + A_{ij}(t),$$

where the indicator variable I_{ikl} is 1 if SKU j has attribute level l in salient attribute k and is 0 otherwise. For example, consumer i's intrinsic value for SKU j with attribute-level combination "Ben & Jerry's, 16 oz., vanilla" is a sum of the cumulative attractions for "Ben & Jerry's," "16 oz.," and "vanilla," and the product j as a whole. The cumulative attractions for an attribute level and a product are updated over time as follows:

$$(3)\qquad A_{ikl}(t) = \phi_k \times A_{ikl}(t-1) + R_{ikl}(t),$$

$$(4)\qquad A_{ij}(t) = \phi_p \times A_{ij}(t-1) + R_{ij}(t),$$

where ϕ_k and ϕ_p are decay factors. The consumer- and attribute-level-specific variable $R_{ikl}(t)$ is the incremental reinforcement consumer i derives from level l in attribute k in time epoch t. Similarly, $R_{ij}(t)$ is the incremental reinforcement consumer i derives from product j as a whole. The difference between the two incremental reinforcements is that the attribute-level reinforcement affects the intrinsic value of all products that share similar attribute levels whereas the product-level reinforcement does not. This distinction captures a product's unique and shared characteristics with other products.

Attribute-Level Reinforcement

The incremental reinforcement for an attribute level l in time epoch t depends on whether it was chosen in time epoch t – 1. If attribute level l were chosen, it would have both consumption and shopping experiences. Otherwise, the incremental reinforcement would have only the shopping experience. The consumption experience occurred in time epoch t – 1 (i.e., before the current store visit), $C_{ikl}(t-1)$, and the shopping experience happened in time epoch t (i.e., during the current store visit), $S_{ikl}(t)$. Shopping experience applies only to the available attribute levels. That is,

$$(5)\qquad R_{ikl}(t) = \begin{cases} C_{ikl}(t-1) + S_{ikl}(t), & \text{if level } l \text{ in attribute k was chosen in } t-1, \\ S_{ikl}(t) & \text{otherwise,} \end{cases}$$

where both consumption and shopping experiences depend on the familiarity with attribute level l. This attribute-level familiarity is a function of the number of previous consumptions. We denote consumer i's familiarity with attribute level l after consumption in time epoch t by $F_{ikl}(t)$. The consumption experience for the attribute level chosen in time epoch t – 1 is given by

$$(6)\qquad C_{ikl}(t-1) = C_{k0} + C_{k1} \times F_{ikl}(t-2).$$

Note that the consumption experience lags behind the shopping experience by one time period, because we define the beginning of the time epoch by the initiation of a store visit. The previous specification has two implications. First, the incremental reinforcements derived from trial-and-repeat consumption are different. Note that $F_{ikl}(t-2) = 0$ if i consumes attribute level l for the first time in time epoch t – 1. Thus, C_{k0} can be interpreted as the reinforcement received from trial consumption. Second, if $C_{k1} < 0$, each subsequent consumption counts less and less as consumer i becomes more familiar with the attribute level l. This captures satiation and implies diminishing marginal utility at the attribute level. In contrast, if $C_{k1} > 0$, there is increasing marginal utility; that is, the consumer likes the attribute level more after each consumption.

Similarly, the shopping experience for an attribute level during time epoch t is given by

$$(7)\qquad S_{ikl}(t) = S_{k0} + S_{k1} \times F_{ikl}(t-1).$$

The shopping experience enables us to make better use of information contained in the available but unchosen attribute levels. Because the set of unchosen attribute levels is large, a modeler could potentially do better by distinguishing the levels according to the level of familiarity the consumer has with each (previous models ignore this information and assume that all unchosen attribute levels have zero shopping experience). Behaviorally, shopping experience captures the intuition that the consumer considers only a small set of attribute levels when purchasing a product, and we hypothesize that the small set of attribute levels includes those with which the consumer is familiar. Thus, it captures how the consumer activates "memory" of attribute levels during the act of choosing (Alba, Hutchinson, and Lynch 1991; Lichtenstein and Srull 1987).

A basic premise of our model is that attribute-level familiarity leads to ease of memory recall. The number of prior consumptions seems a good proxy for measuring familiarity in the absence of other, more direct memory-based measures. We measure familiarity with an attribute level and a product by the number of times they are consumed. For example, if $T_{ikl}(t)$ is the number of times consumer i consumes attribute level l of salient attribute k before and including time epoch t, different functional forms can be posited to relate the familiarity function ($F_{ikl}[t]$) with the

[3]We also estimated the model with i.i.d. double exponential errors. The results are available from the authors on request.

number of consumptions ($T_{ikl}[t]$).[4] We capture a diminishing effect of additional consumption by using a log functional form as follows:

$$F_{ikl}(t) = \ln[1 + \theta_a \times T_{ikl}(t)], \quad (8)$$

where θ_a is a parameter that controls the rate of diminishing for each additional consumption.

Product-Level Reinforcement

Unlike attribute-level reinforcement, product-level reinforcement captures the consumer's idiosyncratic liking for a product beyond shared attribute levels. Consequently, product-level reinforcement for a product affects only its own cumulative attraction and does not influence the attractions of other products. The product-level incremental reinforcement depends on whether product j was chosen in time epoch t – 1. If j were chosen, it would have consumption and shopping experiences; otherwise, it would have only the shopping experience. Again, shopping experience only applies to available products. That is,

$$R_{ij}(t) = \begin{cases} C_{ij}(t-1) + S_{ij}(t), & \text{if product j was chosen in } t-1, \\ S_{ij}(t) & \text{otherwise,} \end{cases} \quad (9)$$

where $F_{ij}(t)$ is consumer i's familiarity with product j after consumption in time epoch t. The consumption experience for the product chosen in time epoch t – 1 is given by

$$C_{ij}(t-1) = C_{p0} + C_{p1} \times F_{ij}(t-2). \quad (10)$$

Analogous to attribute-level consumption, this specification enables us to differentiate between trial and repeat consumption at the product level. If the consumer is new to a product before consumption in time epoch t – 1, we will have $C_{ij}(t-1) = C_{p0}$. If $C_{p1} < 0$, each additional consumption receives a smaller reinforcement because the consumer becomes satiated with the product; if $C_{p1} > 0$, each consumption increases the marginal utility for the next consumption.

Similarly, the shopping experience for product j in time epoch t is given by

$$S_{ij}(t) = S_{p0} + S_{p1} \times F_{ij}(t-1). \quad (11)$$

In a category with many SKUs, the shopping experience singles out a small set of products with which the consumer is familiar. This recognizes that the consumer considers a small set of products before purchase, and the evaluation process consists of memory activation and recall.

As previously, we model product-level familiarity by a log functional form as follows:

$$F_{ij}(t) = \ln[1 + \theta_p \times T_{ij}(t)], \quad (12)$$

where $T_{ij}(t)$ is the number of times consumer i consumes SKU j up to and including time epoch t, and θ_p is the product-level diminishing rate to be estimated.[5]

Our model is related to the GL and FH models in the following ways: The GL model has a product-specific intercept term, whereas the FH model replaces this intercept term with attribute levels the product assumes. Because the total number of attribute levels in all salient attributes is less than the number of SKUs, the FH model uses less parameter. Formally, we specify the GL and FH models as follows:

$$V_{ij}(t) = v_j + A_{ij}(t) \quad \text{(GL model)}, \quad (13)$$

$$V_{ij}(t) = \sum_{k=1}^{K} \sum_{l=1}^{L_k} [v_{kl} + A_{ikl}(t)] \times I_{jkl} \quad \text{(FH model)}, \quad (14)$$

where v_j and v_{kl} are intercept terms associated with product j and level l in salient attribute k, respectively. In addition, their models restrict C_{k1}, C_{p1}, S_{k0}, S_{k1}, S_{p0}, and S_{p1} to be zero. There are no attribute-level or product shopping experiences.

As in the FH model, our model decomposes the intrinsic value of a SKU into its attribute-level components. Thus, our approach yields each consumer's part-worths for all attribute levels at any point in time and can be used to predict the demand for any new product, even when a new attribute level is introduced.[6]

Error Structure

We assume the error structure for utility $U_{ij}(t)$ to include two components: attribute- and product-specific errors. Attribute-specific errors capture serial correlations in attribute-level utilities ($A_{ikl}[t]$), and product-level errors capture serial correlation in product-specific utilities ($A_{ij}[t]$) across time. In particular, we have

$$\varepsilon_{ij}(t) = \sum_{k=1}^{K} \sum_{l=1}^{L_k} \xi_{ikl}(t) \times I_{jkl} + \xi_{ij}(t), \quad (15)$$

where both $\xi_{ikl}(t)$ and $\xi_{ij}(t)$ are assumed to follow an AR(1) autoregressive process of order 1 as follows:

$$\xi_{ikl}(t) = \rho_k \times \xi_{ikl}(t-1) + \upsilon_{ikl}(t), \quad (16)$$

$$\xi_{ij}(t) = \rho_p \times \xi_{ij}(t-1) + \upsilon_{ij}(t), \quad (17)$$

where ρ_k and ρ_p capture the autocorrelations. We assume $\upsilon_{ikl}(t) \sim N(0,\sigma_k^2)$ for all k and l. Similarly, $\upsilon_{ij}(t) \sim N(0,\sigma_p^2)$ for all j. Consequently, $\varepsilon(t)$ is a multivariate normal distribution with means zero and covariance $\Pi(t) = \Sigma_{k=1}^{K} \sigma_k/(1-\rho_k^2) \times I_k'I_k \times \Gamma_k + \sigma_p/(1-\rho_p^2) \times \Gamma_p$, where the (s,t) elements of Γ_k and Γ_p are $\rho_k^{|s-t|}$ and $\rho_p^{|s-t|}$, respectively, and I_k is a matrix of indicator variables for attribute levels of all SKUs. We

[4]In a previous version, we experimented with three different functional forms (step, linear, and log) to capture different rates of memory development: "instantaneous," "constant return to scale," and "diminishing return to scale," respectively. We used Horowitz's (1983) test to determine whether the functional forms have different adjusted pseudo R^2 and found that log function fit the data the best. The log attribute-level familiarity predicts an asymmetric spillover effect on a major brand by priming a minor brand. Nedungadi (1990) finds that priming a minor brand in an unfamiliar attribute level benefits a major brand more than priming it in a familiar attribute level. This is consistent with a log familiarity function, because the marginal increase in familiarity is smaller in a familiar attribute level as a result of priming.

[5]It is reasonable to measure $T_{ikl}(t)$ and $T_{ij}(t)$ over a rolling time horizon to discard distant past consumption experiences and to avoid them increasing indefinitely. We use a one-year time horizon in this article. Krishnamurthi and Raj (1991) define product familiarity $F_{ij}(t)$ as 1 if consumer i chooses j in at least 50% of all previous purchases and as 0 otherwise. Thus, the consumer can be familiar with only one product at a time.

[6]A consumer's familiarity with a new attribute level can be set to zero when the consumer is not aware of the level, and it can be set to a positive value if the consumer is aware of it (perhaps because of advertising or word of mouth).

allow for the same serial correlated error structure for the FH and GL models for ease of comparison.

The Log-Likelihood Function

We control for the effects of the marketing-mix variables with $M_j(t)$ as follows:

$$(18) \qquad M_j(t) = \beta^P \times P_j(t) + \beta^D \times D_j(t) + \beta^{AD} \times AD_j(t).$$

The variables $P_j(t)$, $D_j(t)$, and $AD_j(t)$ are unit price, store display, and advertising feature of SKU j as observed by consumer i during shopping in time epoch t. We use this functional form to benchmark our model against the FH model (which uses the same functional form).

The probability of consumer i choosing SKU j in time epoch t is given by

$$(19) \quad Pr_{ij}(t) = Prob\left[U_{ij}(t) > U_{ij'}(t), \forall j' \neq j, j' \in J(t)\right]$$
$$= \int_{-\infty}^{W_{ij1}(t)} \cdots \int_{-\infty}^{W_{ijJ}(t)} MVN(\eta_{ijj'}, \forall j')\partial\eta_{ij1} \cdots \partial\eta_{ijJ},$$
$$= \Phi\left[W_{ij.}(t)\right],$$

where J(t) corresponds to the set of SKUs available to consumer i on time epoch t. The equation $W_{ijj'} = V_{ij} + M_{ij} - V_{ij'} - M_{ij'}$ corresponds to the difference in deterministic components, and $MVN(\eta_{ijj'}, \forall j')$ is a multivariate normal distribution with $\eta_{ijj'} = \varepsilon_{ij'} - \varepsilon_{ij}$.

Finally, we build in heterogeneity by estimating a two-segment latent-class model (Kamakura and Russell 1989). The log-likelihood function is given as follows:

$$(20) \qquad LL = \sum_i \ln\left\{\sum_{s=1}^{2} \pi^s \times \prod_j \prod_t \Phi\left[W_{ij.}^s(t)\right]^{I_{ij}(t)}\right\},$$

where the indicator variable $I_{ij}(t)$ is 1 if consumer i bought SKU j in time epoch t and is 0 otherwise. The size of the segment s is π^s.

Note that there is neither a product-specific nor an attribute-specific intercept term in our model, which helps reduce the number of parameters. Altogether, there are 5 parameters associated with the update of the cumulative attraction of attribute levels in each of the three salient attributes (i.e., ϕ_k, C_{k0}, C_{k1}, S_{k0}, S_{k1}) and product (i.e., ϕ_p, C_{p0}, C_{p1}, S_{p0}, S_{p1}); 3 marketing-mix, response-sensitivity parameters; and 2 parameters (i.e., θ_a and θ_p) for modeling familiarity for the log model. To identify the model, we set S_{k0} and S_{p0} equal to 1. There are two parameters associated with each of the four error components (ρ_k, ρ_p and σ_k, σ_p). Thus, our model has a total of $5 + 6 \times (K + 1)$ for a product category that has K salient attributes. For K = 3, there is a total of 29 parameters for a single-segment model. For a two-segment model, the number of parameters becomes $2 \times 29 + 1 = 59$.

Behavioral Premises and Rationale

Our model is based on three behavioral premises: (1) the consumer accumulates attribute- and product-level experiences, (2) the experiences include consumption and shopping components, and (3) both consumption and shopping experiences depend on familiarity. These behavioral premises rely on existing research found in consumer behavior and experimental economics.

The first premise is based on the theoretical framework that Lynch, Marmorstein, and Weigold (1988) propose. They suggest that the consumer uses recalled prior attribute- and product-level experiences as input in choosing products. They show that the relative importance of the two kinds of memory recall depends on their relative accessibility and diagnosticity. We believe our model is a first step toward operationalizing this framework in scanner-data research. Our parameters S_{k1} and S_{p1} measure the relative importance of attribute- and product-level familiarity. We can interpret the parameters as their relative diagnosticity for product choice because they translate familiarity into reinforcement and choice probability. Our log functional form parameters θ_a and θ_p transform the number of consumptions into familiarity and, consequently, can be interpreted as their relative accessibility. The higher the θ, the higher the relative accessibility is.

The second premise suggests that the consumer acquires a shopping experience for an attribute level and a product without consuming it by means of mental simulation of whether it would have been better. Camerer and Ho (1999) and Camerer, Ho, and Chong (2002) show that people care about the forgone payoffs of available actions they did not choose but could have chosen. This forgone payoff, which they call simulated reinforcement, is found to be substantial and useful in predicting subjects' switching behavior in strategic games. The shopping experience seems particularly relevant when the consumer is faced with a large number of attribute levels and is likely to have a different shopping experience for each level. For example, if the consumer pays attention only to familiar attribute levels, the shopping experience for those levels is likely to be much more intense than it is for unfamiliar attribute levels. Ignoring shopping experience implies that the consumer treats all unchosen attribute levels identically, which seems unreasonable when the number of attribute levels is large.

The third behavioral premise posits that familiarity is the main determinant of the level of consumption and shopping experiences. Erdem (1998) shows that a consumer's attraction for a product changes as the consumer learns more about the product's attributes through additional uses. Similarly, Alba, Hutchinson, and Lynch (1991) provide three reasons attribute- and product-level familiarity might play a central role in grocery shopping. First, because the grocery-shopping environment is highly complex, consumers often rely on recall to recognize products on the shelf and evaluate them. Second, when consumers look at the grocery store display without preconceptions, attribute-level and product familiarity likely influence how easily specific products catch their eye and enter into their consideration sets. Third, consumers often have very low motivation to spend time when they shop for groceries. For example, Dickson and Sawyer (1986) report that consumers who shop for toothpaste, margarine, coffee, and cold cereal spent an average of 12 seconds from the time they approached the shelf to the time they placed the selected item in their carts. Product- and attribute-level familiarities play a central role in the identification and evaluation of products in this highly efficient shopping process.

Familiarity-based shopping and consumption experiences provide a natural way to account for variety-seeking behavior. This intertemporal switching behavior is well documented (e.g., Bawa 1990; Feinberg 1997; Feinberg, Kahn, and McAlister 1994; Givon 1984; Lattin and McAlister 1985; Trivedi, Bass, and Rao 1994). For a comprehensive review on variety-seeking behavior, see Kahn (1998). A consumer seeks variety if the conditional probability of choosing a product on a given occasion (given that they chose the same product in the last occasion) is lower than the unconditional probability of choosing the product (Kahn, Kalwani, and Morrison 1986). Several researchers have attempted to capture variety seeking by having a negative incremental reinforcement for the chosen attribute levels and product (see, e.g., Lattin 1987; Papatla and Krishnamurthi 1992). In our framework, we model variety-seeking behavior by having $R_{ikl}(t) < 0$ for the chosen attribute level and $R_{ij}(t) < 0$ for the chosen product. This implies that the consumer is less likely to choose the chosen attribute level (product) on the next purchase occasion.[7]

In our model, variety seeking occurs because the consumer becomes satiated with the chosen attribute level and/or receives a simulated reinforcement from unchosen attribute levels that make him or her want to switch to them. We can demonstrate this with Equations 5, 6, and 7. Consider two attribute levels l (chosen) and l' (unchosen). Assume that the latter is twice as familiar as the former (e.g., $F_{ikl'}[t-1] = 2 > F_{ikl}[t-1] = 1$) and $F_{ikl}(t-2) = .5$. Consequently, the incremental reinforcement for the unchosen attribute level l ($R_{ikl'}$) is greater than that for the chosen attribute level (R_{ikl}) if $S_{k1} > C_{k0} + .5 \times C_{k1}$.[8] This results in l' more likely to be chosen than l.

Our model predicts that the consumer will switch to those attribute levels that receive a higher simulated reinforcement from shopping, whereas existing approaches do not make such a prediction. Switching to an unfamiliar attribute level can occur if the products with that attribute level are on promotion (i.e., higher $M_j[t]$ value). Switching back to the familiar attribute levels, which occurs frequently in our data set, is captured by this familiarity-based shopping experience.

EMPIRICAL ANALYSIS

We estimated our model using the method of simulated maximum likelihood. We used the Geweke-Hajivassiliou-Keane recursive probability simulator to evaluate the SKU choice probability in Equation 19 (for details, see Geweke, Keane, and Runkle 1997). We implemented the estimation in a GAUSS program and used dual stopping criteria. We terminated the optimization routine if the changes in parameter estimates were less than 10^{-3} and the improvement in average log-likelihood per observation was less than 10^{-5}.

Data Description

We estimated our model on two Information Resources Inc. scanner-panel data sets, drawn from two different U.S. cities, that capture 133,492 purchase incidences and span 16 product categories. The first data set contains shopping information of 548 households over a two-year period (June 1991–June 1993). It contains purchase information of 15 product categories at five stores located in the same area. These 15 products include 10 food categories (bacon, cola, eggs, frozen pizza, hot dogs, ice cream, potato chips, regular cereal, spaghetti sauce, and yogurt) and 5 nonfood categories (bathroom tissue, detergent, paper towels, soap, and toothpaste). The data set also contains weekly information about product availability at each store and marketing-mix information, such as prices, advertising features, and in-store displays.[9] The second data set captures fabric softener purchases of 594 households over a two-and-a-half-year period (January 1990–June 1992) in Philadelphia. Similar to the first data set, it contains product availability and marketing-mix information. This data set enables us to check the robustness of our model, because the fabric softener has four (rather than three) salient attributes and a different set of panelists who live in a different city and shop over a different time horizon. Table 1 provides detailed information for each product category. We sorted the categories in the total number of attribute levels in all salient attributes.

We defined the input variables for our autoregressive probit model as follows: We computed the price of each SKU according to the price per basic unit (e.g., price per ounce). In addition, we treated the variables ADj(t) (advertising feature) and Dj(t) (in-store display) as zero-one variables.

Unlike the fabric softener data set, the first data set uses three data fields to describe a product: brand name, package size, and flavor. Consequently, we used these three salient attributes to represent the SKUs. We used the product descriptions provided by the manufacturers to delineate attribute levels. To be comprehensive, we treated each different description as a new attribute level. Table 1 gives some examples of brands, package sizes, and flavors for each category. In our data set, two SKUs rarely share the same attribute-level combination. Note that even if two SKUs have the same attribute levels, the consumer could develop different attractions for them because of having a different product-specific experience for each level.

There are an average of 202 SKUs per category. Of the 16 product categories in our data sets, only 3 (bacon, eggs, and fabric softener) have less than 100 SKUs. There are an average of 28 brands, 32 package sizes, and 35 flavors in a category. The soap category has the highest number of brands, 47, and the eggs category has the least number, 12. The frozen pizza category has as many as 145 package sizes, and the eggs category has only 3. In terms of flavor, the ice

[7]Our model does not distinguish between individual-level variety-seeking behavior and intrahousehold taste variation. That is, the model cannot separate purchase behavior of a multiple-member household whose members have different tastes from that of a single-member household who seeks variety as a result of satiation. This is a limitation imposed on our model by data availability. If intrahousehold consumption histories are known, each household member could have a separate consumption experience. However, we can study such distinction by adding an interaction term that captures the size of the household to the consumption and shopping experience.

[8]Different consumers with the same set of parameters will have different variety-seeking propensity if they have different familiarities with the attribute levels. For example, if the familiarity for the unchosen level is 1 rather than 2, the incremental reinforcement for the chosen level is higher than for the unchosen level. These consumers will have a smaller variety-seeking propensity.

[9]Because we cannot detect temporary intraweek stock-outs, we assumed the products experiencing intraweek stock-outs were available throughout the week when we updated the shopping experiences.

cream category has the most, 145 flavors, and the eggs category has 5 different egg sizes.

Estimation Results

Our two-segment model has a fixed 59 parameters for all categories. A two-segment GL or FH model can have hundreds of parameters. For example, a two-segment GL and FH model has 853 and 401 parameters, respectively, for the ice cream category. We benchmarked our model against the GL and FH models for product categories in which these models have fewer than 200 parameters. These "small product categories" (see Table 1) are eggs, fabric softener, bathroom tissue, cola, bacon, paper towels, and hot dogs. We did this for two reasons. First, it is difficult to obtain reliable parameter estimates when a model has hundreds of parameters. Second, we wanted to give our model a stringent test because we developed some of our model constructs (e.g., shopping experience) specifically for "large product categories" that have many SKUs and attribute levels. Because the competing models ignore these model constructs, they are more likely to work well in small product categories.

To estimate our model parameters and validate our model out-of-sample, we divided the 104 weeks of data for all 16 categories except the FH model's fabric softener as follows: We used the first 13 weeks of data for initialization, the next 65 weeks for calibration, and the last 26 weeks for model validation. Fader and Hardie (1996) use 52 weeks of data for initialization in their fabric softener data set, to which we adhered for ease of comparison. We also estimated our model using 7 weeks of initialization period. The two sets of results were not different; we report the 13-week results because prior studies often use at least a three-month of initialization period.[10] A detailed breakdown of the sample size in calibration and validation for all 16 categories is provided in Table 1.

The top half of Table 2 shows the calibration results for small product categories. We use log-likelihood, average hit probability, and adjusted pseudo R^2 to evaluate the

[10]Across categories, the average differences in total log-likelihood over the same time horizon are .5% (22 log-likelihood points) and 1.4% (19 log-likelihood points) in-sample and out-of-sample, respectively. The differences between the two sets of parameter estimates (C_{k1}, C_{p1}, S_{k1}, S_{p1}, ρ_k, σ_k, ρ_p, σ_p) were not statistically significant. Details are available on request from the authors.

Table 2

CALIBRATION AND VALIDATION RESULTS FOR THE SMALL PRODUCT CATEGORIES

	Eggs	*Fabric Softener*	*Bathroom Tissue*	*Cola*	*Bacon*	*Paper Towels*	*Hot Dogs*
Calibration							
Sample size	6252	4417	9303	9241	2383	7768	2577
Log-likelihood							
Our model	–5414	–2600	–11,287	–10,592	–3272	–8845	–3635
Fader and Hardie (1996)	–5699	–3074	–13,196	–11,911	–3523	–9407	–3927
Guadagni and Little (1983)	–5978	–3039	—	—	–3892	—	—
Empirical frequency	–8691	–15,504	–30,384	–34,861	–6000	–25,104	–8502
Average Hit Probability							
Our model	.55	.83	.51	.60	.37	.56	.47
Fader and Hardie	.53	.82	.45	.55	.32	.52	.44
Guadagni and Little	.53	.81	—	—	.27	—	—
Empirical frequency	.33	.03	.06	.03	.12	.05	.06
Adjusted ρ^2							
Our model	.37	.83	.63	.69	.44	.65	.57
Fader and Hardie	.34	.80	.56	.65	.39	.62	.52
Guadagni and Little	.30	.80	—	—	.33	—	—
Validation							
Sample size	2494	2137	3510	3495	842	2889	927
Log-likelihood							
Our model	–2262	–1484	–4357	–3910	–1201	–3194	–1445
Fader and Hardie	–2486	–1814	–5346	–4527	–1283	–3467	–1556
Guadagni and Little	–2518	–1650	—	—	–1521	—	—
Empirical frequency	–3781	–7867	–12,108	–12,463	–2461	–11,781	–3089
Average Hit Probability							
Our model	.56	.81	.50	.58	.39	.57	.46
Fader and Hardie	.53	.80	.42	.53	.33	.54	.42
Guadagni and Little	.55	.79	—	—	.26	—	—
Empirical frequency	.30	.03	.05	.03	.11	.04	.06
Adjusted ρ^2							
Our model	.39	.80	.64	.68	.49	.72	.51
Fader and Hardie	.32	.76	.55	.63	.43	.69	.44
Guadagni and Little	.31	.77	—	—	.33	—	—

Notes: FH's fabric softener has four attributes; thus, the number of parameters are adjusted accordingly. In addition, they initialized with 52 weeks of data, which we adhere to here.

models.[11] Overall, our model performed better than the FH model, which in turn did better than the GL model, in all three measures. The average ρ^2 for the FH model is .55, and the average ρ^2 for our model is .60. Of the seven categories, the category with the smallest improvement (fabric softener) had an adjusted ρ^2 improvement of 4%, whereas the best-performing category (bacon) showed a 13% improvement. The bottom half of Table 2 shows that our model is consistently better in all categories in the validation phase.[12] It predicts an average of an 11% improvement over the FH model in adjusted ρ^2. Our worst-performing category (fabric softener) had a 6% improvement, and our best-performing category (eggs) showed a 20% improvement.

An intuitive way to judge the models is by determining their average hit probability. Table 2 shows that the empirical frequency model and the FH model have an average hit probability of .10 and .52, respectively, in calibration. Our model shows an average hit probability of .56. This represents an average improvement of 10%. A similar pattern occurs in the validation phase. Our model has an average hit probability of .55, and the FH model has an average hit probability of .51.

Table 3 shows the calibration and validation results for our model in nine large categories. The superiority of our model over the empirical frequency model is even more pronounced here than it is in small product categories. In terms of adjusted ρ^2, our model for large (small) categories averages .76 (.60) and .82 (.60) in-sample and out-of-sample, respectively. Note that the improvement over the empirical frequency model is higher out-of-sample than it is in-sample, which suggests that our model does not overfit the data in large product categories. We believe (as evidenced by the results in Table 4) that this can partly be accounted for by a more pronounced shopping experience in large product categories.

Tests of Key Behavioral Premises

Our first behavioral premise posits that the consumer accumulates both attribute- and product-level reinforcements. We can easily test whether this is true by estimating two special cases of the general model: (1) a model without attribute-level reinforcement (i.e., $C_{k0} = C_{k1} = S_{k1} = 0$) and (2) a model without product-level reinforcement (i.e., $C_{p0} = C_{p1} = S_{p1} = 0$). The top panel (labeled "Behavioral Premise 1") of the estimation results (in Table 4) shows that both kinds of reinforcement are necessary for developing a predictive model of SKU choice. In all product categories, both special cases are strongly rejected in favor of the more general model. These results provide a nice, albeit indirect, support of Lynch, Marmorstein, and Weigold's (1988) theoretical framework in the field.

Our second behavioral premise suggests that the consumer accumulates both consumption and shopping experiences. We can test this premise by determining the model

[11]The adjusted pseudo R^2 (ρ^2) measures the proportion of the log-likelihood of the empirical frequency model explained by the model of interest. The adjusted ρ^2 for a model M is given by

$$\rho^2(M) = \frac{LL(0) - LL(M) - NP(M)}{LL(0)}, \tag{21}$$

where NP(M) is the number of parameters for model M, LL(0) is the log-likelihood value of the empirical frequency model, and LL(M) is the maximized log-likelihood value of model M. The empirical frequency model assigns to each product a choice probability based on its aggregate market share in the first 78 weeks of the data set, which cover the initialization and calibration periods. The adjusted ρ^2 is a good measure because it captures the fit while adjusting for the number of parameters.

[12]The GL and FH models cannot predict purchases made to new products or new products that introduce new attribute levels to the category. To give these models their best chance, we excluded these purchases in the validation phase.

Table 3

CALIBRATION AND VALIDATION RESULTS FOR THE LARGE PRODUCT CATEGORIES

	Potato Chips	*Yogurt*	*Spaghetti Sauce*	*Soap*	*Toothpaste*	*Detergent*	*Regular Cereal*	*Ice Cream*	*Frozen Pizza*
Calibration									
Sample size	4395	7949	2701	3197	1892	4596	8262	4351	3396
Log-Likelihood									
Our model	–5485	–6930	–3076	–3605	–1762	–4287	–8998	–2854	–2496
Empirical frequency	–17,868	–36,341	–10,601	–13,007	–7636	–20,008	–36,301	–19,318	–14,915
Average Hit Probability									
Our model	.57	.74	.60	.65	.70	.71	.68	.77	.76
Empirical frequency	.03	.01	.03	.02	.02	.02	.02	.02	.01
Adjusted ρ^2									
Our model	.69	.81	.70	.72	.76	.78	.75	.85	.83
Validation									
Sample size	1698	3189	1085	1268	792	1677	3040	1623	1412
Log-likelihood									
Our model	–1309	–2755	–1147	–1269	–693	–1036	–2686	–880	–894
Empirical frequency	–7210	–15,202	–4304	–5598	–4067	–10,015	–13,265	–7960	–7017
Average Hit Probability									
Our model	.65	.73	.60	.67	.72	.79	.71	.80	.77
Empirical frequency	.04	.01	.03	.02	.02	.02	.02	.02	.01
Adjusted ρ^2									
Our model	.81	.81	.72	.76	.82	.89	.79	.88	.86

Table 4
TESTS OF BEHAVIORAL PREMISES: LOG-LIKELIHOOD RATIOS OF NESTED MODELS

	Small Product Categories							*Large Product Categories*								
	Eggs	*Fabric Softener*	*Bathroom Tissue*	*Cola*	*Bacon*	*Paper Towels*	*Hot Dogs*	*Potato Chips*	*Yogurt*	*Spaghetti Sauce*	*Soap*	*Toothpaste*	*Detergent*	*Regular Cereal*	*Ice Cream*	*Frozen Pizza*
The Full Model (Log-Likelihood)	–5414	–2600	–11,287	–10,592	–3272	–8845	–3635	–5485	–6930	–3076	–3605	–1762	–4287	–8998	–2854	–2496
Behavioral Premise 1																
No attribute-level reinforcement	246*	722*	2370*	936*	568*	1281*	405*	3014*	414*	2110*	1479*	1932*	695*	581*	728*	2598*
No product-level reinforcement	1080*	176*	3868*	3385*	568*	3699*	1013*	4523*	694*	3400*	846*	2196*	618*	2215*	293*	2748*
Behavioral Premise 2																
No consumption experience	306*	70*	737*	326*	237*	1487*	370*	3226*	1299*	1646*	234*	2202*	309*	230*	44*	533*
No shopping experience	775*	1027*	2566*	1725*	556*	1751*	449*	5134*	724*	3288*	1801*	2168*	2093*	8038*	1680*	1621*
Behavioral Premise 3																
No familiarity	765*	549*	3636*	1498*	1623*	3075*	719*	1763*	696*	2259*	3469*	2074*	955*	4813*	293*	645*

*$p < .01$.

fits of two special cases of the model: (1) a model without consumption experience (i.e., $C_{k0} = C_{k1} = C_{p0} = C_{p1} = 0$) and (2) a model without shopping experience (i.e., $S_{k1} = S_{p1} = 0$). The middle panel of the results (labeled "Behavioral Premise 2") strongly suggests that both kinds of experiences are crucial in fitting and predicting SKU choice. Note that the likelihood ratios are much higher in large product categories. This is indicative of the greater importance of shopping experience in these categories.

Our third premise is that both shopping and consumption experiences depend on familiarity. Thus, the consumer can have a different incremental reinforcement for the same attribute level or product over time. We test this premise by estimating a special case of our model, where $C_{k1} = S_{k1} = C_{p1} = S_{p1} = 0$. The bottom panel of estimation results (labeled "Behavioral Premise 3") strongly suggests that both consumption and shopping experiences are familiarity based. The results suggest that consumers use memory cues to narrow down products during shopping and to derive marginally greater utility in repeated consumptions.

Shopping and Consumption Experiences

Our extensive data set enabled us to develop several empirical regularities on shopping and consumption experiences across categories. This effort is exploratory in nature and focuses on the unique features of our model (i.e., S_{k1}, C_{k1}, S_{p1}, and C_{p1}).

As we discuss subsequently, positive S_{k1} and S_{p1} suggests a memory activation effect that results in switching to familiar attribute levels or products. A negative C_{k1} and C_{p1} indicates decreasing marginal utility, whereas a positive C_{k1} and C_{p1} imply increasing marginal utility at the attribute and product levels. Table 5 shows the parameter estimates.

There is a "memory activation" effect for familiar attribute levels. Of the 70 statistically significant S_{k1}s, 60 are positive. Thus, in general, shopping experience increases with attribute-level familiarity. This occurs in at least one segment of all 14 significant categories for brand, 14 of 15 significant categories for size, and 12 of 14 significant categories for flavor. This finding suggests that when consumers switch away from the chosen attribute levels, they are more likely to switch to familiar attribute levels. This propensity to choose familiar attribute levels (e.g., brand) is consistent with Erdem and Keane's (1996) finding that risk-averse consumers avoid less familiar brands because they are uncertain about their benefit.

Product-level shopping experience also increases with familiarity. The 30 S_{p1} parameters are significant and positive in all categories. Therefore, consumers tend to switch back to products with which they are familiar. This implies that consumers are reluctant to spend time evaluating unfamiliar products during shopping.

We observed an increasing marginal utility at the attribute level in a majority of the categories. Of the 50 statistically significant C_{k1}s, 40 are positive. This phenomenon occurs in at least one segment in 10 of 11 categories with significant C_{k1} for brand, 8 of 10 categories for size, and 12 of 14 categories for flavor. However, the same phenomenon occurs less frequently at the product level. Only 17 of the 30 C_{p1}s are statistically different from zero; 12 of them are positive, which suggests increasing marginal utility at the product level.

The memory for attribute consumption is less "accessible" and less "diagnostic" than the memory for product consumption. We find that $\theta_p > \theta_a$ in at least one segment of 11 of 16 product categories, and $S_{p1} > S_{k1}$ in at least one segment of all 16 categories. Because the parameters θ_a and θ_p convert the number of consumptions into familiarity, a higher θ implies better accessibility to the memory of past consumptions. However, the relative diagnosticity of these familiarities in consumer choice depends on the values of the parameters S_{k1} and S_{p1}, because they translate these familiarities into reinforcements and attractions. In short, these results suggest that past product consumption is easier to recall than is past attribute consumption, and remembered product consumption influences consumer choice more heavily than does remembered attribute consumption.

Autoregressive Error Structures

The parameters ρ_k and ρ_p enable us to study whether there is any serial correlation in random utilities over time. Table 6 shows the parameter estimates for ρ_k and ρ_p. A majority of the attribute-level correlation parameters (61 of 98) are not significantly different from zero. Of those that are significantly different from zero, only 8 have an absolute value greater than .5. Half of these higher serial correlations occur in brand. A majority of the product-level serial correlation parameters (22 of 32) are not significantly different from zero. Of those that are significantly different from zero, only 4 have an absolute value greater than .5. They occur in the bacon, cola, and bathroom tissue categories. Overall, our results suggest that serial correlation in random utilities over time is modest.

The variances of the attribute-level error terms (σ_k) provide a clue as to the degree of correlation among products that share similar attribute levels. Most of the estimated σ_ks are small (compared with the product-level variance term σ_p; see Table 6) except for bathroom tissue, bacon, hot dogs, toothpaste, and soap. These results suggest that correlations among utilities of SKUs are small within a product category at a particular purchase incidence.

DISCUSSION AND CONCLUSION

In this research, we show that it is possible to develop an SKU choice model with parameters that are independent of the number of SKUs and the number of attribute levels of a product category. With three salient attributes, our model has only 59 parameters (compared with an average of 199 and 118 for the GL and FH models for seven small product categories). Our model uses all data to describe and predict choices made to all SKUs. We also show that this highly parsimonious model performs substantially better in log-likelihood, average hit probability, and adjusted pseudo R^2. This superior performance occurs in all product categories. In addition, we demonstrate that our model describes and predicts choices well in nine large product categories.

We developed our model by modifying the standard utility specification and by incorporating familiarity-based consumption and shopping experience at both the attribute and the product levels. Our results suggest that both the attribute-level and the product-level familiarities are important for predicting SKU choice in small and large product categories. If familiarities for attribute levels and products are induced by a consumer's memory for them, our results

Table 5

MAXIMUM LIKELIHOOD PARAMETER ESTIMATES OF SHOPPING AND CONSUMPTION EXPERIENCES

	Small Product Categories							*Large Product Categories*								
	Eggs	*Fabric Softener*	*Bathroom Tissue*	*Cola*	*Bacon*	*Paper Towels*	*Hot Dogs*	*Potato Chips*	*Yogurt*	*Spaghetti Sauce*	*Soap*	*Toothpaste*	*Detergent*	*Regular Cereal*	*Ice Cream*	*Frozen Pizza*
Segment 1 Parameters																
Shopping Experience, S_{kl}																
Brand	.22*	–.02	.29*	.26*	.01	.35*	.40*	.02	–.01	.15*	.49*	.16*	.12*	.17*	.03	.29*
Size	.08*	.33*	.16*	.35*	.23*	.05*	.05*	.01	.10*	–.04	.03	.21*	.27*	–.15*	–.07*	.16*
Flavor	–.02	.02	.06*	.54*	.40*	.11*	.32*	.04*	.21*	–.14*	.46*	.29*	.06	–.02	.64*	.14*
Formula		.14*														
SKU, S_{pl}	.60*	.32*	.27*	.40*	.48*	.34*	.95*	.53*	.83*	.79*	.86*	.54*	.47*	.96*	.34*	–.38*
Consumption Experience, C_{kl}																
Brand	.12*	.00	.11*	.07*	–.01	.19*	–.02	.02	–.02	.21*	–.14*	–.19*	.04	–.01	.65*	.18*
Size	.01	.00	.08*	.05*	.05*	.06*	.17*	–.01	.09*	.04	.03	–.27*	.04	.03*	.00	.06*
Flavor	.09*	.00	.09*	.04*	.11*	.02	–.01	–.11*	.00	.10*	–.03	.25*	.07*	.12*	–.16*	–.34*
Formula		.03*														
SKU, C_{pl}	–.02	.01	.27*	.05*	.06*	.08*	.10*	–.07*	–.02	–.01	.14*	–.08*	.44*	.08*	.00	.00
θ_a	5.11*	15.06*	4.27*	2.80*	3.87*	4.61*	5.49*	4.89*	4.79*	4.94*	6.34*	5.09*	5.17*	1.58*	4.72*	4.64*
θ_p	5.15*	.74*	4.27*	4.73*	6.59*	4.57*	1.54*	1.09*	.85*	4.97*	5.17*	4.70*	5.10*	5.37*	4.78*	4.77*
Segment 2 Parameters																
Shopping Experience, S_{kl}																
Brand	.10*	.30*	–.25*	.22*	1.10*	.21*	.14*	–.02	.00	.00	–.56*	–.01	.12*	.21*	.43*	.12
Size	–.04	.13*	.20*	.12*	.24*	.14*	.15*	.07*	.00	.42*	.43*	–.19*	–.09*	–.32*	.27*	.21*
Flavor	–.05*	.00	.11*	.15*	.27*	–.05	.24*	.10	.00	–.12*	.67*	.11	–.02	.05*	.02	.33*
Formula		.16*														
SKU, S_{pl}	.46*	.51*	1.04*	.48*	.50*	.57*	.45*	.31*	1.03*	.39*	.77*	.93*	.31*	.70*	.77*	.59*
Consumption Experience, C_{kl}																
Brand	.04*	.03	.00	.05*	–.03	.15*	.16*	–.14*	.00	–.04	.28*	.41*	.18*	.02	.19	.22*
Size	.04	.04*	.04*	.08*	.16*	–.02	–.26*	–.07*	.00	.03	–.09*	–.39*	.17*	.03*	.11	–.06
Flavor	.07*	.07*	.04*	.00	.14*	–.09*	.13*	.17*	.00	.11*	.07*	–.17*	.09*	.06*	–.02	–.02
Formula		.01														
SKU, C_{pl}	–.04*	.01	–.02	.02	.38*	.14*	.13*	–.15*	–.14*	–.02	.01	.25*	.03	.06*	.14	.06
θ_a	4.77*	15.06*	5.05*	2.41*	4.96*	4.72*	4.96*	5.24*	5.08*	4.38*	4.29*	4.32*	4.75*	1.14*	4.99*	4.96*
θ_p	4.82*	1.01*	4.20*	5.36*	5.14*	4.95*	.43*	2.41*	2.64*	4.81*	5.51*	4.85*	4.86*	4.90*	5.09*	5.01*

*$p < .01$.

Table 6

MAXIMUM LIKELIHOOD PARAMETER ESTIMATES OF AUTOREGRESSIVE ERROR STRUCTURES

	Small Product Categories							*Large Product Categories*								
	Eggs	*Fabric Softener*	*Bathroom Tissue*	*Cola*	*Bacon*	*Paper Towels*	*Hot Dogs*	*Potato Chips*	*Yogurt*	*Spaghetti Sauce*	*Soap*	*Toothpaste*	*Detergent*	*Regular Cereal*	*Ice Cream*	*Frozen Pizza*
Segment 1 Parameters																
Serial Correlation, ρ_k																
Brand	.00	.00	.52*	.19*	–.10*	–.32*	–.01	.00	.00	.00	.68*	–.36*	.00	.00	.00	.00
Size	.00	.00	–.35*	–.07*	.31*	–.06*	.75*	.00	.00	.00	.07*	.05	.00	.00	.00	.00
Flavor	.00	.00	.59*	–.32*	–.06*	–.17*	–.18*	.00	.00	.00	.00	–.19*	.00	.00	.00	.00
Formula		.00														
SKU, ρ_p	.00	.00	.61*	.75*	–.18*	.21*	.22*	.00	.00	.00	.00	–.22*	.00	.00	.00	.00
Variance, σ_k																
Brand	.49*	.05*	.60*	.15*	1.73*	.80*	.67*	.41*	.02	.06	.17*	.47*	.01	.08*	.03	.28*
Size	.14*	.01	.25*	.53*	1.13*	.06*	.54*	.27*	.00	.04	.42*	.24*	.34*	.05*	.04*	.25*
Flavor	.17*	.04*	.49*	.16*	4.59*	.23*	.49*	.28*	.01	.07*	.15*	.44*	.26*	.04*	.14*	.41*
Formula		.04*														
SKU, σ_p	1.79*	1.65*	2.02*	1.42*	2.91*	1.97*	2.11*	1.92*	1.64*	1.78*	1.62*	1.72*	1.68*	1.70*	1.80*	1.89*
Segment 2 Parameters																
Serial Correlation, ρ_k																
Brand	.00	.00	.64*	.00	–.31*	.69*	.11*	.00	.00	–.23*	–.29*	.46*	.00	.00	.00	.00
Size	.00	.00	–.15*	.00	.17*	.15*	.30*	.00	.00	.10*	.36*	.22*	.00	.00	.00	.00
Flavor	.00	.00	–.18*	.00	.20*	.29*	–.66*	.00	.00	–.23*	.52*	.09	.00	.00	.00	.00
Formula		.00														
SKU, ρ_p	.00	.00	–.62*	.00	.54*	.14*	.02	.00	.00	.00	–.09*	.04	.00	.00	.00	.00
Variance, σ_k																
Brand	.00	.11*	.61*	.10*	1.09*	.12*	.95*	.23*	.02	.11*	4.14*	.18*	.10*	.04*	.01	.09*
Size	.10*	.07*	3.14*	.12*	2.39*	.22*	2.33*	.17*	.01	1.19*	.40*	.34*	.32*	.03*	.01	.10*
Flavor	.85*	.01	.69*	.17*	3.31*	.33*	1.01*	.01	.05*	.02	2.20*	4.92*	.18*	.06*	.03	.10
Formula		.02														
SKU, σ_p	1.72*	1.64*	2.24*	1.75*	2.51*	1.76*	1.79*	1.76*	1.69*	1.68*	2.16*	1.77*	1.72*	1.68*	1.64*	1.66*

$^*p < .01$.

Table 5

support the notion of memory-based decision making (Alba, Hutchinson, and Lynch 1991). In some way, we have shown that incorporating research findings from consumer research can be a powerful method for improving the descriptive and predictive power of a choice model in the scanner-data literature.

We modeled attribute- and product-level familiarities as a function of the number of consumptions in the respective attribute levels and products. We are quick to point out that attribute- and product-level familiarities can also be a function of other factors, such as television commercials, word-of-mouth communication, and consumer reports. For example, Erdem and Keane (1996) use advertising exposure, besides the number of consumptions, to model brand familiarity.[13] It would be worthwhile to explore how these other factors affect attribute-level and product familiarity in the future.

We have found strong evidence that shopping experience exists at both the attribute and the product levels and increases with familiarity. To the best of our knowledge, this is the first demonstration of the effect of shopping experience on product choice in the scanner-data literature. This result suggests that the consumer uses attribute- and product-level familiarities to narrow down product alternatives during shopping. This enables us to capture the frequent phenomenon that the consumer may occasionally experiment with a new product but often returns to buying the existing set of familiar products. The "memory activation" effect provides a theoretical rationale for the occurrence of variety-seeking behavior that is frequently observed in our product categories.

There are immediate and future effects of price and nonprice promotion. Our notion of shopping experience provides a behavioral mechanism by which the future benefit of promotion can be realized. If promotion leads to higher familiarity, and higher familiarity leads to increased shopping experience, promotion can increase future product purchases. Because product-level shopping experience tends to be stronger and easier to recall than attribute-level (e.g., brand) shopping experience, managers might find it more effective to engage in product-level promotion than in attribute-level promotion.

An alternative way to interpret shopping experience is to examine the exploitation of information contained in unchosen attribute levels and products. Unlike other models, our models do not treat all unchosen attribute levels and products equally. We assume the consumer pays special attention to those attribute levels and products that have been consumed on previous occasions. However, our extraction of information from the unchosen attribute levels and products is somewhat simplified. More sophisticated approaches, particularly behavioral-based approaches, can be formulated to differentiate attribute levels and products to obtain a better fit and prediction. For example, consumers may be allowed to have imperfect memory and may gradually forget what they have bought previously. This will lead to a different familiarity function that may improve fit and prediction power. We suggest this subject for further research.

Finally, we suggest a few ways the proposed model can be used by brand managers and is currently used in practice. The first is base volume forecasting. Our model can be used to forecast regular sales volume (i.e., base volume) of any SKU in a product category. Our model reveals the relative contribution of each attribute level to the base volume while controlling for marketing-mix effects. The second is relative importance of each attribute. When using the model, the relative importance of each salient attribute can be easily analyzed. This analysis can be done at the individual consumer level and across time. The third is forecasting sales for line extensions. As indicated previously, an attractive feature of our model is its ability to forecast sales for line extensions, despite whether they introduce new attribute levels to a category.

[13]Erdem and Keane (1996) use the commercial viewing file of a household to determine the advertising exposure of a brand. Unfortunately, we do not have similar information at the SKU level to enhance the way we model attribute- and product-level familiarities.

REFERENCES

Alba, Joseph W., J. Wesley Hutchinson, and John G. Lynch Jr. (1991), "Memory and Decision Making" in *Handbook of Consumer Behavior*, T. Robertson and H. Kassarjian, eds. Englewood Cliffs, NJ: Prentice Hall, 1–49.

Allenby, Greg M. and Peter E. Rossi (1991), "Quality Perceptions and Asymmetric Switching Between Brands," *Marketing Science*, 10 (3), 185–204.

Bawa, Kapil (1990), "Modeling Inertia and Variety Seeking Tendencies in Brand Choice Behavior," *Marketing Science*, 9 (3), 263–78.

Ben-Akiva, Moshe and Steven R. Lerman (1985), *Discrete Choice Analysis*. Cambridge: Massachusetts Institute of Technology Press.

Bucklin, Randolph E. and Sunil Gupta (1992), "Brand Choice, Purchase Incidence, and Segmentation: An Integrated Modeling Approach," *Journal of Marketing Research*, 29 (May), 201–215.

Camerer, Colin and Teck-Hua Ho (1999), "Experience-Weighted Attraction Learning in Normal-Form Games," *Econometrica*, 67 (4), 837–74.

———, ———, and Juin-Kuan Chong (2002), "Sophisticated EWA Learning and Strategic Teaching in Repeated Games," *Journal of Economic Theory*, 104 (1), 137–88.

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

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

Chong, Juin-Kuan, Teck-Hua Ho, and Christopher Tang (2001), "A Modeling Framework for Category Assortment Planning," *Manufacturing and Service Operations Management*, 3 (3), 191–210.

Dickson, P. and A. Sawyer (1986), "Point-of-Purchase Behavior and Price Perceptions of Supermarket Shoppers," Working Paper No. 86–102, Marketing Science Institute, Cambridge, Massachusetts.

Erdem, Tülin (1998), "An Empirical Analysis of Umbrella Branding," *Journal of Marketing Research*, 35 (August), 339–51.

——— and Michael P. Keane (1996), "Decision-Making Under Uncertainty: Capturing Dynamic Brand Choice Processes in Turbulent Consumer Goods Markets," *Marketing Science*, 15 (1), 1–20.

Fader, Peter S. and Bruce G.S. Hardie (1996), "Modeling Consumer Choice Among SKUs," *Journal of Marketing Research*, 33 (November), 442–52.

———, James M. Lattin, and John D.C. Little (1992), "Estimating Nonlinear Parameters in the Multinomial Logit Model," *Marketing Science*, 11 (4), 372–85.

Feinberg, Fred M. (1997), "A Dynamic Inertial Brand Choice Model Incorporating Interbrand Similarity Structure," working paper, Faculty of Management, University of Toronto.

———, Barbara E. Kahn, and Leigh McAlister (1994), "Implications and Relative Fit of Several First-Order Markov Models of Consumer Variety-Seeking," *European Journal of Operational Research*, 76 (2), 309–320.

Geweke, John F., Michael P. Keane, and David E. Runkle (1997), "Statistical Inference in the Multinomial Multiperiod Probit Model," *Journal of Econometrics*, 80 (1), 125–65.

Givon, Moshe M. (1984), "Variety Seeking Through Brand Switching," *Marketing Science*, 3 (Winter), 1–22.

Guadagni, Peter M. and John D.C. Little (1983), "A Logit Model of Brand Choice Calibrated on Scanner Data," *Marketing Science*, 2 (3), 203–238.

Ho, Teck-Hua and Christopher Tang (1998), *Research Advances in Product Variety Management*. Boston: Kluwer Academic Publishers.

Horowitz, J. (1983), "Statistical Comparison of Non-Nested Probabilistic Discrete Choice Models," *Transportation Science*, 17 (2), 319–50.

Kahn, Barbara E. (1998), "Variety: From the Consumer's Perspective," in *Research Advances in Variety Management*, Teck-Hua Ho and Christopher S. Tang, eds. Boston: Kluwer Academic Publishers, 19–37.

———, Manohar U. Kalwani, and Donald G. Morrison (1986), "Measuring Variety-Seeking and Reinforcement Behaviors Using Panel Data," *Journal of Marketing Research*, 23 (May), 89–100.

Kamakura, Wagner A. and Gary J. Russell (1989), "A Probabilistic Choice Model for Market Segmentation and Elasticity Structure," *Journal of Marketing Research*, 26 (November), 379–90.

Krishnamurthi, L. and S.P. Raj (1991), "An Empirical Analysis of the Relationship Between Brand Loyalty and Consumer Price Elasticity," *Marketing Science*, 10 (2), 172–83.

Lattin, James M. (1987), "A Model of Balanced Choice Behavior," *Marketing Science*, 6 (Winter), 48–65.

——— and Leigh McAlister (1985), "Using a Variety-Seeking Model to Identify Substitute and Complementary Relationships Among Competing Products," *Journal of Marketing Research*, 22 (August), 330–39.

Lichtenstein, Meryl and Thomas K. Srull (1987), "Processing Objectives as Determinants of the Relationship Between Recall and Judgment," *Journal of Experimental Social Psychology*, 23 (March), 93–118.

Lynch, John G., Jr., Howard Marmorstein, and Michael F. Weigold (1988), "Choices from Sets Including Remembered Brands: Use of Recalled Attributes and Prior Overall Experiences," *Journal of Consumer Research*, 15 (September), 225–33.

Manski, C. and S. Lerman (1977), "The Estimation of Choice Probabilities from Choice-Based Samples," *Econometrica*, 45 (8), 1977–88.

Nedungadi, Prakash (1990), "Recall and Consumer Consideration Sets: Influencing Choice Without Altering Brand Evaluations," *Journal of Consumer Research*, 17 (3), 263–76.

Papatla, Purushottam and Lakshman Krishnamurthi (1992), "A Probit Model of Choice Dynamics," *Marketing Science*, 11 (2), 189–206.

Quelch, J. and D. Kenny (1994), "Extend Profits, Not the Product Lines," *Harvard Business Review*, 72 (September–October), 153–60.

Siddarth, S., Randolph E. Bucklin, and Donald G. Morrison (1995), "Making the Cut: Modeling and Analyzing Choice-Set Restriction in Scanner Panel Data," *Journal of Marketing Research*, 32 (August), 255–66.

Trivedi, Minakshi, Frank M. Bass, and Ram C. Rao (1994), "A Model of Stochastic Variety-Seeking," *Marketing Science*, 13 (3), 274–97.