PETER S. FADER and BRUCE G. S. HARDIE*

Most choice models in marketing implicitly assume that the fundamental unit of analysis is the brand. In reality, however, many more of the decisions made by consumers, manufacturers, and retailers occur at the level of the stock-keeping unit (SKU). The authors address a variety of issues involved in defining and using SKUs in a choice model, as well as the unique benefits that arise from doing so. They discuss how a set of discrete attributes (e.g., brand name, package size, type) can be used to characterize a large set of SKUs in a parsimonious manner. They postulate that consumers do not form preferences for each individual SKU, per se, but instead evaluate the underlying attributes that describe each item. The model is shown to be substantially superior to a more traditional framework that does not emphasize the complete use of SKU attribute information. Their analysis also highlights several other benefits associated with the proposed modeling approach, such as the ability to forecast sales for imitative line extensions that enter the market in a future period. Other implications and extensions also are discussed.

Modeling Consumer Choice Among SKUs

To most contemporary choice modelers, the fundamental unit of analysis is the brand. Entering the supermarket, however, the brand becomes a product line comprising many stock-keeping units (SKUs). Consumers typically choose among SKUs on the basis of a set of product attributes, which tend to be discrete and tangible. For example, when buying toothpaste, Yoshinuri normally chooses a 4.6-ounce tube of mint-flavored tartar control Crest gel. However, when motivated by promotions, he will sometimes choose another item. His resulting choice reflects not only his preference for the promoted brand but also his preference for package size and type (e.g., medium-sized tubes), product form (gel), formula (tartar control), and flavor (mint). If Yoshinuri has a strong preference for gel formulations, he is unlikely to switch to a brand that does not offer a gel-based SKU.

Thus, in contrast to most choice models presented in the marketing literature, brand choice is rarely a final decision by itself; rather, SKU choice is a more fitting description of the overall decision process. Brand is obviously an important component of this decision, but the other components cannot be ignored.

The importance of SKUs in today's marketplace has attracted a great deal of attention in the business press. For example, as part of its move toward efficient consumer response, Procter & Gamble sought to reduce its total number of SKUs by approximately 25% (Schiller 1993). Many of these reductions were achieved by trimming product lines (e.g., eliminating slow-selling detergent package sizes), but several brands (e.g., White Cloud and Charmin bathroom tissue) were consolidated as well.

With such significant changes presently taking place in many markets, it is important that managers understand how consumers evaluate and integrate different SKU attributes in making choice decisions. Nevertheless, most choice modelers in marketing typically assume away most of these critical details. It is ironic that these aspects are so vital to the decisions made by consumers, manufacturers, and retailers, yet receive so little attention in formal choice models. This irony is further magnified because these descriptive characteristics are readily available in most scanner panel data sets.

We address this gap between current choice models and marketing practice. Our primary goal is to describe the direct benefits of modeling consumer choice among SKUs. We discuss how today's standard models deal with this issue and then demonstrate a powerful yet parsimonious modeling approach that enables a researcher to include all of the distinguishing attributes that characterize a particular product category's set of SKUs. In developing our model and our database for empirical testing, several other interesting issues emerge, including (1) the ability to include many more

*Peter S. Fader is Associate Professor of Marketing, the Wharton School, University of Pennsylvania. Bruce G. S. Hardie is Assistant Professor of Marketing, London Business School. The authors thank John Walsh and Len Lodish for their extraordinary efforts in developing this article and appreciate extensive comments from Bart Bronnenberg, Bob Meyer, the Editor, one anonymous *JMR* reviewer, and numerous participants from seminars presented at the following schools: Massachusetts Institute of Technology, University of Wisconsin (Madison), New York University, Cornell University, Penn State University, Washington University (St. Louis), University of Texas (Austin), Temple University, and Wharton. The data used here come from Information Resources, Inc., and the authors acknowledge the efforts of Lionel Wang for his help in constructing the data set.

choice alternatives (i.e., dozens of SKUs instead of the 3–10 "brands" that most academic articles use) and (2) the ability to produce forecasts for new line extensions that are not present during the model calibration period. We discuss several other benefits as well.

The article proceeds as follows. First, we discuss the role of attributes in consumer choice settings and review how previous choice models have incorporated (or ignored) SKU attribute information. Second, we motivate our general modeling approach and describe the specification to be used here. Third, this model and several benchmarks are tested on a highly realistic fabric softener data set, which includes 56 unique SKUs. Fourth, we apply these models to the problem of forecasting the sales of two new SKUs that were launched after the end of the model calibration period. Fifth, and finally, we conclude with a discussion of several issues that arise as a consequence of our modeling approach and identify several areas worthy of further research.

OVERVIEW OF SKU ATTRIBUTES

For most consumer packaged goods, it is convenient to describe SKUs in terms of a set of categorical attributes, that is, physical characteristics that uniquely identify every one of the items available on the store shelf. For example, the scenario mentioned at the outset of the article uses six different SKU attributes (brand, package size and type, product form, formula, and flavor) to portray the toothpaste market.[1] Several of these attributes, especially brand and package size, are common to virtually all product categories. Others vary from one category to another. Beyond the attributes, we use the term *attribute level* to denote each of the distinct offerings that together constitute an attribute; for example, Crest, Colgate, and Aquafresh are a subset of the levels under the attribute *brand* in the toothpaste category.

Marketing research firms such as Information Resources, Inc. (IRI) and A. C. Nielsen use several criteria to determine what can be treated as an SKU attribute. First, the attribute must be *consumer recognizable*. That is, an SKU attribute must be easily observable by casually examining the front of the package. Second, SKU attributes also should be *objective*: There should be no ambiguity about each item's precise level for each SKU attribute. For example, a suitable SKU attribute for bottled water might be fruit flavoring (e.g., unflavored versus orange flavored versus lemon flavored). But this attribute should generally measure the presence of these flavorings, not the intensity of the flavoring. Third, looking across all the items available in a product category, an SKU attribute must be *collectively exhaustive*: Every SKU attribute must apply to every SKU (though the levels may differ from item to item). Many SKU attributes are measured simply as the presence or absence of a particular product characteristic; this leads to two levels on the defining attribute. For example, even if only a small number of spaghetti sauce brands have "homestyle" varieties, all of the other items would be represented by a "not homestyle" or "regular" attribute level.

With this in mind, it is easy to see how the number of SKU attributes and levels can explode to an unmanageable size for a complex product category. Much of the analyst's skill, therefore, is to define a minimal number of distinct attributes that capture all of the relevant traits within the product category. For example, if "homestyle" applies to only one brand in the category, it might be viewed as an irrelevant SKU attribute.[2] The key objective is that each SKU should represent a unique combination of attributes. Once this is achieved, then additional attributes may be redundant.

To recapitulate, the analyst's challenge is to define a set of SKU attributes with dual competing objectives in mind: (1) a desire to capture all the richness of the SKU-by-SKU variability within the product category and (2) a desire to use as few SKU attributes as necessary to do so. In managerial practice, the typical number of SKU attributes ranges between three and eight, though few academic models have considered more than two attributes. Similarly, though few academics have used more than four to six levels for a given attribute, realistic product categories frequently have twenty or more levels for attributes such as flavor (e.g., in categories such as soup and cat food). Another attribute with many levels is brand, which has ten or more distinct entries within many categories.

The Use of SKU Attributes in Choice Models

Given their pioneering use of scanner data for choice modeling, it is no surprise that Guadagni and Little (1983; hereafter G&L) were among the first to acknowledge the role of SKU attributes: "In applying the model, the alternatives will be products, but their exact level of aggregation and which ones to include in the relevant set are not necessarily obvious. Should different flavors, or colors, or sizes be treated as different products or lumped together?" (G&L, p. 211). Guadagni and Little did not answer this question completely, but they did at least recognize the importance of package sizes as well as brand names. Although brand loyalty was the single strongest explanatory variable in their model, G&L found that size loyalty was the second strongest predictor, showing greater statistical significance than any other marketing mix effect.

Despite this precedent, relatively few researchers have used additional SKU attributes in choice models; most tend to suppress all SKU attributes besides brand. Their approach is to either pool items across an attribute by creating a composite average (e.g., Pedrick and Zufryden 1991), treat items with any different SKU characteristics as distinct brands (e.g., Fader and Lattin 1993), or vary the competitive set on the basis of the attribute levels of the chosen item. For example, Kamakura and Russell (1989) observe the package size of the item actually chosen and then drop all items with different package sizes in calculating purchase probabilities at that time. We next introduce our proposed modeling approach, which obviates the need to use any of these shortcuts.

[1]We exclude from this initial discussion nonproduct attributes associated with different marketing mix activities (e.g., price, promotional indicators), though such variables will be incorporated into the full models, which we implement subsequently.

[2]The "goodness" of our attribute-based model is driven by the commonality of the attribute levels across the brands in the category. If each brand has many unique attributes or levels, this model may not be appropriate. Fortunately, the majority of attributes and levels in most categories are shared by several brands.

MODEL DEVELOPMENT

The standard approach to modeling product choice with scanner data involves the use of the familiar multinomial logit model (MNL), which has the following structure:

$$p_i = \frac{\exp(v_i)}{\sum_{i'} \exp(v_{i'})},$$

where, ignoring household and time indices, p_i is the probability of choosing item i, and following McFadden (1973), v_i is called the deterministic component of item i's utility. The v_i are viewed as having two components: (1) a preference component, $v_{pref}(i)$, which represents the household's base preference toward item i, and (2) a marketing mix component, $v_{mm}(i)$, which represents the effects of marketing variables (e.g., price, presence or type of promotion) on the household's choice behavior.

Over the past decade, much research effort has focused on modeling heterogeneity in $v_{pref}(i)$. Guadagni and Little specify it by using a homogeneous SKU-specific constant, α_i, augmented by exponentially smoothed loyalty variables to model each household's unique preference over time. Other researchers (e.g., Krishnamurthi and Raj 1991; Lattin and Bucklin 1989) have created alternative measures of loyalty using other purchase feedback-based variables; in all cases, these have been combined with α_i to convey the household's core preference toward each item.

Kamakura and Russell (1989) develop an alternative approach to capturing cross-sectional heterogeneity in the preference component of v_i by using a MNL mixture model. They assume the existence of a finite number of homogeneous segments in the market, with probability of segment membership being invariant across households. The v_i are therefore conditional on segment membership (i.e., $v_i \mid s$). Each segment's preference component is captured by a vector α_s, where α_{is} represents the s^{th} segment's preference toward item i. Other approaches to capturing heterogeneity in $v_{pref}(i)$ include the use of parametric or semiparametric random effects methods (Chintagunta, Jain, and Vilcassim 1991) or Dirichlet-like models (Fader and Lattin 1993).

Regardless of the specific approach taken to modeling heterogeneity in $v_{pref}(i)$, a common aspect is that they contain SKU-specific intercept terms, α_i. The presence of these intercepts becomes problematic when these models are applied to relatively large data sets. Consider the set of 56 fabric softener SKUs described in Table 1. A standard logit model would require us to estimate 55 α_i's (in addition to marketing mix effects). A mixture model with two segments would require 111+ parameters, and a three-segment model would need 157+ parameters. With this explosion in the number of parameters, such models quickly become infeasible. The problem is even worse when using a semiparametric (Chintagunta, Jain, and Vilcassim 1991) approach—a model with three support points would require 279+ parameters. Clearly there is a problem—we identify the need to model consumer choice at the SKU level, but existing modeling techniques are not well-suited for the task.

Thus, before resolving this issue of how to accommodate consumer heterogeneity, it is essential to consider the item-specific intercepts (α_i) more carefully. In using these dummy variables, the researcher implicitly assumes that consumers maintain preferences for each individual SKU. Returning to the toothpaste scenario, this implies that Yoshinuri has a *direct* preference for 4.6-ounce tubes of mint-flavored tartar control Crest gel, as well as for each of the other 100 to 150 SKUs typically available in the toothpaste category. On further thought, however, it is clear that consumers do not retain such a broad set of well-formed preferences. More realistically, Yoshinuri's preference for 4.6-ounce tubes of mint-flavored tartar control Crest gel is likely to be derived from his preferences over each of the individual SKU attributes. This suggests that, rather than modeling the direct preferences for each SKU, we should model the preferences over the levels of each SKU attribute.[3]

There is a long tradition of modeling consumer preferences among multiattribute alternatives as a function of the consumer's preferences over the attributes that describe the choice alternatives. The theoretical justification for such an approach can be found in psychology (Fishbein 1967) and economics (Lancaster 1971; Quandt 1956). Moreover, in both the theoretical (e.g., Anderson, de Palma, and Thisse 1992) and empirical (e.g., Berry 1994) literature on discrete choice models of product differentiation, it is common to posit that a consumer's utility for a product is not a direct function of the product itself, but instead depends on the product's attributes and the consumer's tastes.

This tradition suggests an alternative approach to specifying the preference component of v_i: Rather than express it as a function of SKU-specific intercept terms (i.e., $v_{pref}(i) = f(\alpha_i)$), which implicitly assumes that consumers maintain preferences toward each individual SKU, we model consumer preferences over the attributes that describe the SKUs in the product category (and only indirectly over the SKUs themselves). In other words, the preference component of v_i can be expressed as

$$v_{pref}(i) = g(l_1, l_2, \ldots, l_N),$$

where the N-tuple $(l_1, l_2, \ldots, l_N)$ represents the unique set of attribute levels associated with SKU i.[4] In both the marketing and economics literature, it is common to assume an additive utility function that implies (assuming no interactions)

$$g(l_1, l_2, \ldots, l_N), = \sum_{n=1}^{N} g_n(l_n),$$

where $g_n(l_n)$ is the attribute-specific preference function for attribute n.

In addition to its strong theoretical base, this approach to specifying the preference component of v_i offers other benefits as well. Consider the complete set of SKUs outlined in

[3]Here, brand name is considered as one of several SKU attributes, and individual brand preferences are estimated in the same way as preferences toward different package sizes, product forms, and so on. A natural parallel is the inclusion of brand name variables in a conjoint analysis. Although it is possible to decompose these brand name effects (e.g., Chintagunta 1993; Kamakura and Russell 1993) or estimate them as the component of overall preference not explained by the other attributes (e.g., Srinivasan 1979), this is not the objective of our research. Additionally, we contend that when making in-store SKU choice decisions, the consumer is not evaluating each SKU's brand name on multiple dimensions; rather, he or she is using some overall evaluation of each brand, which we interpret as his or her preference for each brand.

[4]Let the set of SKUs in a given product category be defined over N attributes and let L_n be the number of levels associated with the n^{th} attribute. We can represent each SKU by the N-tuple $(l_1, l_2, \ldots, l_N)$, where l_j can take on integer values $1, \ldots, L_j$.

Table 1
SKUs AND THEIR ATTRIBUTES

SKU #	*Brand*	*Form*	*Formula*	*Size*	*Sales*
1	Arm & Hammer	Sheets	Regular	Small	11
2	Arm & Hammer	Sheets	Regular	Medium	75
3	Arm & Hammer	Sheets	Regular	Large	11
4	Bounce	Sheets	Regular	Medium	86
5	Bounce	Sheets	Regular	Extra Large	12
6	Bounce	Sheets	Staingard	Small	63
7	Bounce	Sheets	Staingard	Medium	60
8	Bounce	Sheets	Staingard	Large	17
9	Bounce	Sheets	Unscented	Medium	43
10	Cling Free	Sheets	Regular	Small	111
11	Cling Free	Sheets	Regular	Medium	49
12	Downy	Concentrated	Regular	Small	21
13	Downy	Concentrated	Regular	Small	35
14	Downy	Concentrated	Regular	Medium	128
15	Downy	Concentrated	Regular	Large	58
16	Downy	Concentrated	Regular	Extra Large	22
17	Downy	Concentrated	Light	Small	11
18	Downy	Concentrated	Light	Medium	76
19	Downy	Concentrated	Light	Large	30
20	Downy	Concentrated	Light	Extra Large	5
21	Downy	Refill	Regular	Small	338
22	Downy	Refill	Regular	Medium	2
23	Downy	Refill	Light	Small	267
24	Downy	Sheets	Regular	Small	86
25	Downy	Sheets	Regular	Medium	125
26	Downy	Sheets	Light	Medium	122
27	Final Touch	Concentrated	Regular	Small	32
28	Final Touch	Concentrated	Regular	Medium	241
29	Final Touch	Concentrated	Regular	Large	49
30	Generic	Concentrated	Regular	Extra Large	91
31	Generic	Concentrated	Regular	Extra Large	56
32	Generic	Sheets	Unscented	Large	55
33	Private Label	Concentrated	Regular	Small	23
34	Private Label	Concentrated	Regular	Medium	46
35	Private Label	Concentrated	Regular	Extra Large	5
36	Private Label	Liquid	Regular	Medium	20
37	Private Label	Liquid	Regular	Extra Large	33
38	Private Label	Liquid	Regular	Extra Large	68
39	Private Label	Sheets	Regular	Medium	168
40	Private Label	Sheets	Regular	Large	41
41	Private Label	Sheets	Regular	Extra Large	38
42	Private Label	Sheets	Unscented	Medium	80
43	Snuggle	Concentrated	Regular	Small	22
44	Snuggle	Concentrated	Regular	Medium	379
45	Snuggle	Concentrated	Regular	Large	32
46	Snuggle	Concentrated	Regular	Extra Large	20
47	Snuggle	Concentrated	Light	Medium	276
48	Snuggle	Concentrated	Light	Large	43
49	Snuggle	Sheets	Regular	Small	133
50	Snuggle	Sheets	Regular	Medium	77
51	Snuggle	Sheets	Regular	Large	19
52	Snuggle	Sheets	Light	Medium	66
53	Sta-Puf	Concentrated	Regular	Medium	177
54	Sta-Puf	Concentrated	Regular	Large	120
55	Toss n' Soft	Sheets	Regular	Medium	140
56	Toss n' Soft	Sheets	Regular	Large	3

Table 1. These SKUs can be described by using ten levels on the brand attribute, four levels on the size attribute, four levels for form, and four levels for formula. By focusing on the SKU attributes, as opposed to the SKUs themselves, any attempt to capture heterogeneity in the preference component of v_i requires us to model preferences over the 22 attribute levels, rather than the 56 SKUs required by the traditional model.

Parsimony is not the only advantage. As was previously mentioned, such an approach enables us to infer attribute importance and forecast the sales of certain types of new SKUs. Additionally, there is the matter of low-share items. As would be expected with most panel data sets, many of the SKUs have small choice shares: For our fabric softener data set, there are 24 SKUs with less than 1.0% choice share. Clearly, we would not have much confidence in any estimates of α_i for these SKUs. However, we do not suffer from this problem when focusing on the attribute levels. When we aggregate the data in Table 1 for each attribute level, we find that the least common level (the Arm & Hammer brand name) still has a 2.2% share. In sharp contrast, there are only 14 SKUs with choice shares of 2.2% or higher.

Model Specification and Estimation

For this initial implementation, we assume a main-effects only model. As is the case with most linear models, a main-effects only specification explains a large part of the total variance and produces relatively accurate forecasts (Dawes and Corrigan 1974; Johnson, Meyer, and Ghose 1989). Furthermore, it is worth noting that interactions can easily be accommodated within this framework.[5]

We first specify the functional form of the attribute-specific preference functions, $g_n(l_n)$. Given the nominal nature of SKU attributes, we let $g_n(l_n) = m_{in}\alpha_n$, where m_{in} is an elementary (row) vector, the l_n^{th} element of which equals 1, and α_n is the vector of preferences over the L_n levels of attribute n. We can therefore express the preference component of v_i as

$$v_{pref}(i) = \sum_{n=1}^{N} m_{in} \cdot \alpha_n.$$

Heterogeneity in $v_{pref}(i)$ is introduced through α_n. We employ the modified latent class procedure used previously by Bucklin and Gupta (1992) and Chintagunta (1992). This involves the addition of attribute-specific loyalty variables to Kamakura and Russell's (1989) MNL mixture modeling framework. We model the vector of preferences over the L_n levels associated with attribute n using a vector of segment-specific attribute-level intercept terms, α_{0n}^s. (As with SKU-specific intercepts, one of the intercept terms is constrained to zero for each attribute.) To capture additional sources of heterogeneity across consumers, these intercepts are augmented by a set of attribute-specific loyalties for attribute n, $ATTLOY_n$, that is, the exponentially smoothed variable of G&L.[6] We therefore express the vector of preferences over the levels of attribute n as

$$\alpha_n^s = [\alpha_{0n}^s + \alpha_{1n}^s ATTLOY_n]. \quad (1)$$

The resulting definition of the deterministic component of utility is

$$v_i \mid s = \sum_{n=1}^{N} m_{in} \cdot [\alpha_{0n}^s + \alpha_{1n}^s ATTLOY_n] + \beta^s X_i.$$

We subsequently compare this specification with that of a "pure" Kamakura and Russell (1989) model that includes heterogeneous intercepts but none of the $ATTLOY_n$ variables (i.e., $\alpha_{1n}^s = 0 \forall n,s$).

The $\beta^s X_i$ term represents the marketing mix component of the segment-specific deterministic component of utility. The X_i vector includes the values of the marketing variables for SKU i, and β^s is the vector of response coefficients for these variables; by making this segment-specific, we are able to capture heterogeneity in response to marketing variables across segments.

The probability of choosing SKU i, conditioned on membership to preference segment s, is

$$p_i \mid s = \frac{\exp(v_i \mid s)}{\sum_{i'} \exp(v_{i'} \mid s)}.$$

The unconditional probability of choosing SKU i is therefore

$$p_i = \sum_{s=1}^{S} \theta^s p_i \mid s, \quad (2)$$

where θ^s is the probability of belonging to segment s ($0 < \theta^s < 1 \; \forall s = 1, ..., S$ and $\Sigma_s \theta^s = 1$). These θ^s parameters also can be interpreted as the relative size of segment s.

Up to this point, both the choice occasion and household indices have been suppressed; let them be defined as t and h respectively. By reintroducing them, $p_i \mid s$ becomes $p_i^h(t \mid s)$. Let H be the number of households in the panel, T_h the number of choice occasions for household h, and δ_{it}^h a purchase indicator equal to 1 if household h chooses SKU i on purchase occasion t, and 0 otherwise. For the attribute-based model specification with S preference segments, model parameters are estimated by maximizing the log-likelihood function,[7]

$$LL = \sum_{h=1}^{H} \ln\left[\sum_{s=1}^{S} \theta^s \prod_{t=1}^{T_h} \prod_{i} p_i^h(t \mid s)^{\delta_{it}^h}\right].$$

The number of segments is determined by carrying out the estimation for an increasing number of segments until there is no significant improvement in model fit. Early MNL mixture modeling efforts (e.g., Bucklin and Gupta 1992; Kamakura and Russell 1989) used the heuristic of choosing the number of segments to minimize the Akaike Information Criterion or Bayesian Information Criterion. Following the argument of Bozdogan (1987), however, we use the more conservative CAIC (consistent Akaike Information Criterion) measure to assess relative fit: CAIC = $-2LL + k(\log(N) + 1)$, where LL = log-likelihood, k = number of parameters, and N = number of observations.

EMPIRICAL ANALYSIS

We calibrate our models using scanner panel data for fabric softeners. Because this product is sold in multiple formulations, package sizes, and so on, it is well-suited for our proposed analysis. The data come from an IRI panel in Philadelphia and cover the period from January 1990 to June 1992. The only criterion for inclusion in our data set is that the household must have made at least one purchase in 1991. This gives us 594 qualifying households that made a total of 9781 purchases over the 2½-year period. The 3227 purchases from 1990 are used for initializing model variables, and the 4417 purchases from 1991 are used for model calibration. The remaining 2137 purchases made during the first half of 1992 are used for forecasting purposes.

[5]Because of the large number of potential interactions, the process of adding interaction terms should be driven by the analyst's knowledge of the product category. For example, if a brand "owned" an attribute-level for a long time prior to its use by any other brand, we may want to include a brand × attribute-level interaction term in our model to capture any brand-specific associations (cf. Rangaswamy, Burke, and Oliva 1993). For example, in the toothpaste category, the baking soda attribute-level is strongly associated with Arm & Hammer's Dental Care brand, whereas sensitive formulations are strongly associated with the Sensodyne brand name. A complete analysis of the toothpaste category should examine these types of interactions.

[6]Let $attloy_{jn}$ be the j^{th} element of $ATTLOY_n$ ($j = 1, L_n$). Adding time indices, this is defined as $attloy_{jn}(t+1) = \lambda_n attloy_{jn}(t) + (1 - \lambda_n)\Sigma_i \delta_{it}$, where $\delta_{it} = 1$ if SKU i was purchased at t, 0 otherwise, and Σ_i is over all the SKUs possessing level j of attribute n.

[7]Following Kamakura and Russell (1989), θ^s is modeled as $\exp(\tau^s)/\Sigma_{s'}\exp(\tau^{s'})$ to ensure that $0 < \theta^s < 1$.

Beyond the attribute-specific intercepts and loyalty variables discussed in the previous section, our model includes a standard set of marketing variables: regular price, price cut, and binary indicators for retail promotions (displays and newspaper features).

The complete set of SKUs available during 1991 is presented in Table 1. Unlike other scanner panel data-based studies, we use the complete set of SKUs available; we believe this is a necessity if researchers want to address managerially relevant problems with their modeling efforts. Alternatively, if a researcher were to view the data strictly from the brand level, he or she might be inclined to retain only the top four brands (Downy, Snuggle, Private Label, and Final Touch), which collectively have a 73% share. Unfortunately, such a screening process would eliminate 6 of the 16 best-selling SKUs and would skew the coverage of the other attribute levels as well.

It is also interesting to note that all 22 unique SKU attribute levels (ten brands, four sizes, four forms, and four formulas) in this data set appear at least once within the top 24 items in the market. In other words, the 32 slowest-selling items put no demands on the choice model by requiring additional model parameters. In contrast, the traditional MNL model needs an additional 32 SKU-specific intercepts to accommodate these items (which have a combined share of 24%).

Results: Model Performance

We begin by estimating two benchmark models. The first is the "attribute level-specific intercepts-only" model, which is a natural replacement for the common SKU-specific intercepts-only model (i.e., equating p_i to SKU i's market share) used in other studies. The second benchmark is the "standard" G&L model specification with 55 SKU-specific intercept terms (α_i) and an exponentially smoothed loyalty variable for each of the four SKU attributes. Turning our attention to the attribute-based model, we first note that the one-segment model is equivalent to a G&L specification, but with the 55 SKU-specific α_i's replaced by 18 attribute level-specific intercept terms. In terms of straight (i.e., uncorrected) log-likelihood, this model must perform worse than the standard G&L specification, because the former is merely a constrained version of the latter. The value of our proposed modeling approach becomes clear when we introduce multiple latent segments. We keep adding latent segments until the CAIC criterion is minimized. The results are shown in Table 2.

Based on CAIC, we find that the two-segment model provides the best fit to our data set. As dramatic evidence of the power of the attribute-based model, we see that the two-segment model has a better uncorrected log-likelihood than G&L's model *even though it has ten fewer parameters.* It is extremely unusual to outperform the G&L benchmark with a more parsimonious model.

Table 2
CALIBRATION RESULTS

	LL	*Number of Parameters*	*CAIC*
Intercepts Only	−16396	18	32962
Standard G&L	−10844	67	22318
Attribute-Based:			
1 segment	−11116	30	22514
2 segments	−10610	57	21755
3 segments	−10489	84	21767

Parameter Estimates

Rather than bombard the reader with the parameter estimates for every model, we examine only two models, the one- and two-segment attribute-based models. We focus our discussion on the "part-worths" (i.e., contributions to utility) for each of the attribute levels present in the data set.

Previously, we showed that the preference for each attribute is captured in two components: the (homogeneous) intercept terms and the (heterogeneous) loyalty variables. As is implied by Equation 1, we recombine these into a single measure, using the estimated coefficients as weights. These part-worths are heterogeneous across households, but for ease of interpretation we show only their mean values. For the two-segment model, we compute within-segment weighted means; each household's $ATTLOY_n$ vectors are weighted by its posterior probabilities of segment membership.[8]

The attribute-specific preferences for the one-segment model are shown in Figure 1. The most preferred brand, size, form, and formula are, respectively, Sta-Puf, Medium, Refill, and Regular. These estimates are similar to the part-worths in conjoint analysis, so relative importances can be inferred from the breadth of the range between the highest and lowest partworth for each attribute. By this criterion, the most important attribute is brand, and the least important is formula.[9]

The one-segment model provides a reasonable first pass toward understanding the underlying preferences for each product attribute level, but it certainly does not provide the complete picture of consumer preferences. From numerous previous conjoint studies, we know that part-worths may vary across people. In Figure 2, we show the attribute-specific estimates for the two-segment model. The parameters indicating the relative sizes of the segments (θ^1 and θ^2, not shown) suggest that segments 1 and 2 account for 29% and 71% of consumers, respectively. In some ways, the two-segment model merely confirms the findings of the one-segment model, namely, extra large package sizes and unconcentrated liquids are consistently disliked. But in other respects, the two-segment model reveals preference patterns that are masked in the pooled model. The most striking example is that of refills versus sheets: In the one-segment model, both forms have similar preferences, but in the two-segment model, we see a clear division between a group of consumers who prefer refills (segment 1) and those who prefer sheets (segment 2). It is also interesting to note that the relative importance of the form attribute is substantially greater in segment 1 than in segment 2.

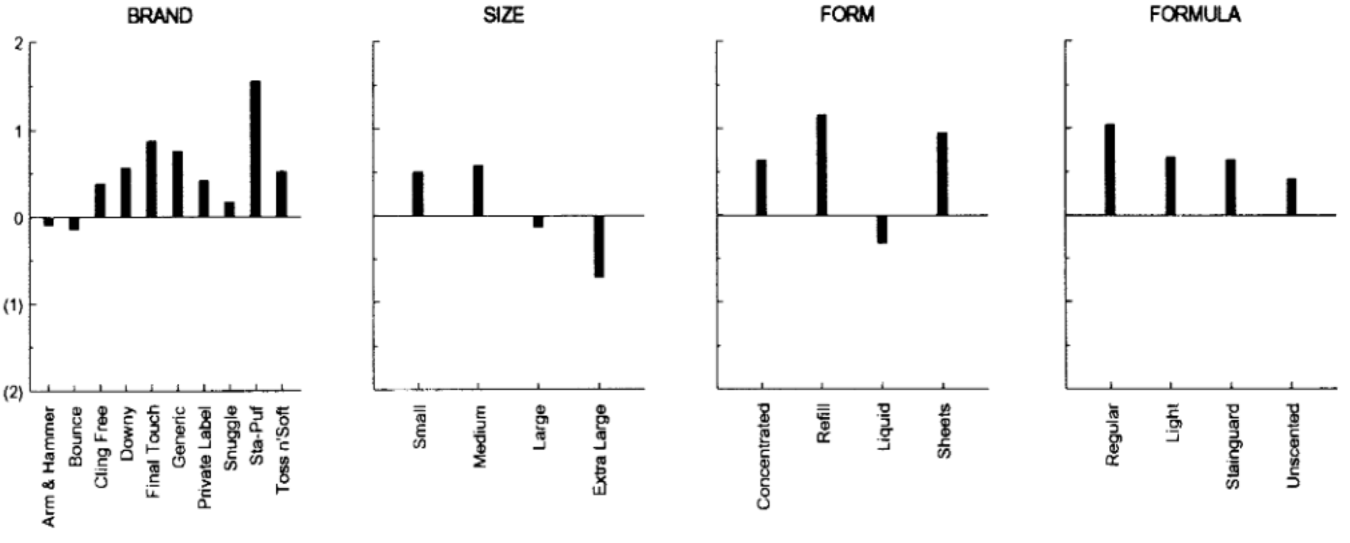
Comparison With an Alternative Model Specification

Although some conceptual and practical benefits arise from including the attribute-specific loyalty variables ($ATTLOY_n$) in our model, there are potential costs as well. Several authors (Lattin 1990; Little and Anderson 1994; Ortmeyer, Lattin, and Montgomery 1991; Russell and

[8]For details on computing posterior probabilities of segment membership, see Kamakura and Russell's (1989, p. 381) study.

[9]Although this is the simplest way to infer attribute importance, there are more complex (and theoretically more appealing) alternative approaches worth considering (see, e.g., Green and Krieger 1995).

Figure 1
ONE-SEGMENT MODEL



Feinberg 1995) have suggested that loyalty variables based solely on previous purchase histories may be "contaminated" by the marketing mix variables and may therefore lead to biased parameter estimates. For example, if a household consistently repurchases a heavily promoted brand, we are unable to determine whether the household is truly loyal to the brand or just sensitive to promotions. If we are to have faith in the validity of the parameter estimates generated by our model, we must determine whether this type of contamination is present in the results just discussed.

Accordingly, we have rerun the three attribute-based models without using any loyalty variables at all (i.e., $\alpha^s_{ln} = 0$ in Equation 1). This alternative model is a direct extension of the original framework in Kamakura and Russell's (1989) study. Our principal focus is to compare the raw intercepts from this model with the loyalty-adjusted estimates discussed previously. If we find a high degree of correspondence between the two sets of parameters, we would have some reassurance that the aforementioned adjustment procedure mitigates the problem of contamination to a large extent. In Table 3, we show the attribute-by-attribute correlations between these two sets of models.

These correlations can be interpreted as follows: Each of the two competing models (i.e., no-loyalty and loyalty-adjusted intercepts) has ten brand-specific terms. The Pearson correlation coefficient between these two sets of estimates is .891 for the one-segment model. Likewise, the four size-specific terms have a correlation of .989 for the one-segment model, and so on. These numbers indicate that the loyalty-adjusted intercepts line up well with the raw intercepts. Casual examination of both sets of estimates in conjunction with the marketing mix variables does not indicate the presence of any biases or systematic relationships with prices or promotions.

Two conclusions can be drawn from these results. First, we see that, in this particular context at least, a proper loyalty adjustment can mitigate many of the possible problems that may arise from trying to interpret the attribute-specific intercept terms in a model with loyalty variables. Second, it is worth noting that the simpler no-loyalty model seems to offer the same substantive implications as the loyalty-based one. This finding may have useful implications for researchers (and managers) who care much more about parameter inference than model fit.[10] As research into the issue of loyalty contamination continues, it will be interesting to see whether these results are typical of the findings that emerge from other data sets.

APPLICATION—FORECASTING SALES OF NEW SKUs

We now examine the forecasting performance of the models developed in the previous section, focusing in particular on the problem of predicting the sales of new SKUs. In many product categories, we see a stream of new SKUs entering the market on a frequent basis. The analyst is faced with a serious problem in trying to handle new items that enter the market after the end of the calibration period. In the absence of SKU-specific constants for these late entrants, it is impossible to provide sales or share forecasts. As a result, no published article using scanner data has ever provided sales or share forecasts for new items that enter during a longitudinal holdout period.

After reflecting on our proposed model, it is apparent that we can provide such forecasts in many cases. Specifically, if a new item is a combination of SKU features that currently exist in the market, we can forecast its sales. Such new SKUs are known as *imitative* or *filling-in* line extensions (Hardie 1994). An imitative line extension occurs when a brand incorporates an attribute level that is new to the brand but already exists in the product category (e.g., Crest baking soda toothpaste, which was the third major brand to introduce baking soda as a feature); a filling-in line extension occurs when a brand produces a new combination of attribute levels that it already possesses. Lest we believe that such

[10]It is important to note, however, that model fit is considerably worse for the no-loyalty specifications. For example, the three-segment no-loyalty model has a log-likelihood of −12795, which is much worse than even the one-segment loyalty-based model.

Figure 2
TWO-SEGMENT MODEL

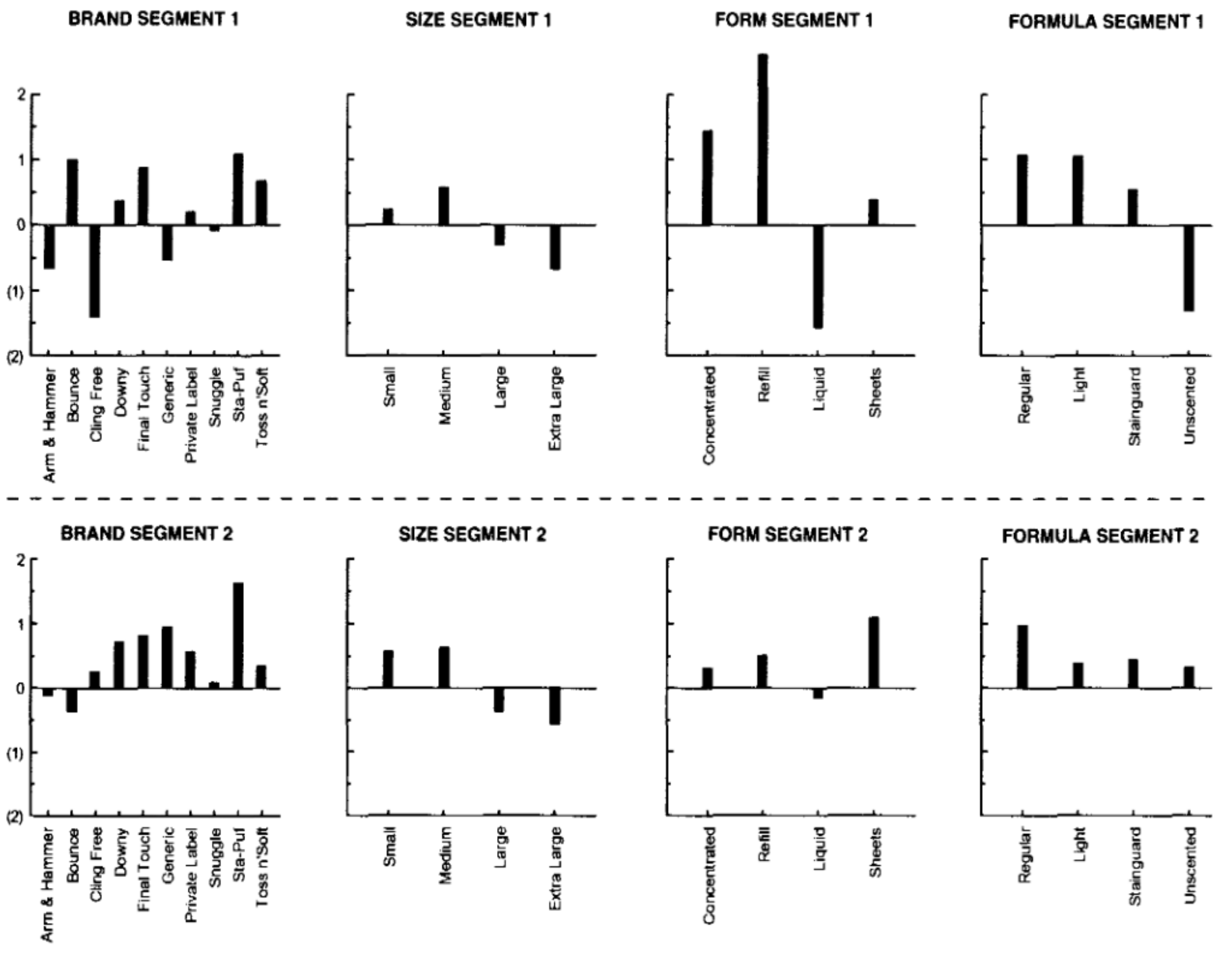


Table 3
CORRELATION OF PARAMETERS FROM LOYALTY AND NO-LOYALTY MODELS

	One-Segment Model	*Two-Segment Model*	
		Segment 1	*Segment 2*
Brand	.891	.849	.586
Size	.989	.874	.961
Form	.941	.504	.998
Formula	.692	.945	.962

types of new products are uncommon, it should be noted that approximately 70% of new consumer packaged goods products are line extensions (Lawrence 1993).

Two assumptions are associated with the use of our methodology for forecasting the performance of certain types of new SKUs. First, we assume that there is no category expansion associated with the launch of the new SKU—the only sources of volume for the line extension are brand shifting and cannibalization. This is not an unrealistic assumption for a relatively mature product category. Second,

Table 4
NEW LINE EXTENSIONS

SKU	*Brand*	*Size*	*Form*	*Formula*
A	Private Label	Medium	Liquid	Light
B	Sta-Puf	Small	Refill	Regular

we assume that preferences over SKU attributes are stable over time. In other words, the presence of a new item using a preexisting attribute level does not change consumers' preference toward that attribute level, per se. Both of these assumptions are rather conservative; any violations would only serve to reduce the accuracy of our share forecasts.

As was noted in the previous section, our fabric softener data set includes a six-month forecast period (January–June 1992). This is a true holdout period, which is not used for model calibration. During this period, two line extension SKUs appeared in the Philadelphia market (see Table 4). Both of these new SKUs are imitative line extensions: SKU A represents the first Private Label light formula SKU, which was an attribute level previously associated with

Table 5
FORECAST RESULTS: LOG-LIKELIHOODS

	Forecast LL	
	Without SKU A & B (n = 2063)	*With SKU A & B (n = 2137)*
Intercepts Only	–8093	–8487
Standard G&L	–5641	—
Attribute-Based:		
1 segment	–5399	–5672
2 segments	–5337*	–5565*
3 segments	–5397	–5730

*Denotes best forecasting model.

Table 6
VOLUME SHARE FORECASTS FOR NEW LINE EXTENSIONS

	Expected Volume Share		*Error Measures*	
	SKU A (act = .26%)	*SKU B (act = 2.66%)*	*MSE*†	*MAPE*††
Intercepts Only	.45	7.85	13.463	135.22
Standard G&L	—[a]	—[a]	—[a]	—[a]
Attribute-Based:				
1 segment	.10	3.11	.111	38.28
2 segment	.11	2.75	.015*	30.70*
3 segment	.08	2.92	.049	39.10

[a]A model with SKU-specific intercepts is not able to generate forecasts for holdout samples that include new SKUs.
†MSE = mean squared error.
††MAPE = mean absolute percentage error.
*Denotes best model under each criterion.

Downy and Snuggle SKUs; SKU B represents the first refill SKU for Sta-Puf, which was an attribute level previously used by Downy.

We compare the relative performance of the various models on the basis of their log-likelihoods for the forecast period. These are calculated both for the complete set of SKUs and for all but the two line extensions (see Table 5). (Only in the latter case can we include the standard G&L model, because it requires SKU-specific constants for all available items.) In Table 5, we show that all the attribute-based models outperform G&L by a wide margin. Additionally, we observe that the two-segment model offers the best fit in the forecast period. This confirms that the higher-order models (i.e., three or more segments) are overfitting the calibration period data, as was suggested by the CAIC measure.

We now examine the volume market share forecasts for the line extension SKUs. Each household's choice probabilities are computed by using a function similar to Equation 2, except that θ^s is replaced by the household's posterior probability of segment membership. Because fabric softeners come in multiple forms (i.e., liquids and sheets), we compute equivalent-units volume, using IRI's conversion numbers. In Table 6, we show that the proposed model provides us with a reasonable approximation of the new items' actual performance. On the basis of both mean-squared error and mean absolute percentage error, the two-segment solution is the clear winner.

Although these forecasts are by no means perfect, even the worst of the forecasts shown in Table 6 is a good first-cut estimate of the new items' actual performance, especially when we consider that no primary data are being used. We do not claim that these results are necessarily representative of what might be seen in other applications. In particular, in situations in which the attribute weights are changing over time, these forecasts could possibly be systematically biased. But under most circumstances, these results offer much encouragement about our ability to use attribute-based methods to forecast the market share of imitative and filling-in line extensions.

DISCUSSION

We argue that the best unit of analysis for most scanner-based choice models is the SKU. Because of the large number of SKUs in most grocery product categories, however, such an approach would place an excessive burden on existing modeling methodologies. Recognizing that SKUs can be described in terms of a small set of discrete attributes and that consumer choice is often made on the basis of these attributes, we propose that researchers model preferences over SKU attributes rather than the SKUs themselves.

We demonstrate that our model represents a powerful yet parsimonious approach to modeling choice at the SKU level. We obtain a set of attribute-specific preference estimates that offer interesting and managerially useful diagnostics about the set of SKUs available in the marketplace. Finally, an additional benefit of this approach is the ability to forecast the performance of different types of line extensions, and our initial results are encouraging in this regard.

Contrasts With Other Modeling Approaches

Several researchers (e.g., Grover and Dillon 1985; Kannan and Wright 1991) have used attribute-based approaches to model hierarchical consumer choice processes. This suggests that a nested logit formulation can be used as an alternative approach for modeling choice among SKUs. However, we argue against such an approach for two reasons. First, as the number of SKU attributes increases, the hierarchical structures become increasingly complex. For example, with our fabric softener data set, there are 24 possible hierarchies (representing every factorial combination of brand, size, form, and formula). If we allow for heterogeneity in hierarchy structure, the modeling effort can quickly become unmanageable. Second, a common motivation for a nested logit formulation is a concern about the IIA assumption associated with the standard MNL model. But the IIA property of our model is a valid assumption, thereby removing this motivation for a nested model.[11]

The reader also may have noted an analogy between our attribute-based models and conjoint analysis. One difference is that our model describes and uses actual products in the marketplace, whereas the stimulus set in a typical conjoint experiment uses artificial objects, many of which do not exist in reality. A second difference is that we explicitly accommodate marketing mix effects, which can have a highly significant impact on consumer choice in many contexts. These covariates generally vary over time and across alternatives and are estimated simultaneously with our attribute-based part-worths, as would be done in an ordinary analysis of covariance model. As a third difference, our choice task (i.e., routine grocery purchasing) is more natural and less obtrusive than that of any laboratory-based conjoint experi-

[11]This is demonstrated in an appendix available from the authors.

ment. This touch of reality makes our model an appealing complement to the standard conjoint methodology. There are interesting possibilities of merging the two methods, especially because conjoint can incorporate new attributes (and levels) that do not currently exist in the market. The development of such a combined model would be a useful extension to this research.

Managerial Implications and Further Research

The motivation for the proposed model stems from issues faced by real-world decision makers, so we clearly have several contributions to offer to practicing managers. We have already highlighted the improvements our model provides in terms of performance and parsimony. The model's other novel benefit—the ability to forecast sales of different types of line extensions—also has significant ramifications in practice, especially because few firms conduct any formal analyses to predict item performance. We show how historical data can be used to analyze the potential performance of certain types of line extensions in a virtually costless manner.

Another managerial benefit is parameter interpretability. In general, it is hard to tell any meaningful stories about the alternative-specific intercepts that emerge from a standard logit model. In contrast, our attribute-specific estimates are readily interpretable and can be highly insightful. For example, line extension opportunities can be identified by comparing attribute-level preferences with the proportion of SKUs in the category possessing each level. A high-preference attribute level that lacks SKU depth could be ripe for line extension.

A natural application of the proposed methodology would be to help fine-tune a product line. How does a firm go about eliminating SKUs in a profitable manner? Is it sufficient to drop the slowest-selling items or is it wiser to eliminate all items sharing an ineffective SKU attribute level (e.g., dropping a particular formula or package size)? Our model lends itself nicely to the type of "what if" simulation analyses frequently employed by IRI and other researchers (Honnold, Brooks, and Little 1990) to address this type of issue.

In closing, we emphasize that this study is intended to be the beginning of a more comprehensive research program on the topic of SKU choice. We hope that our model and qualitative observations encourage others to examine this neglected area from various different perspectives.

REFERENCES

Anderson, Simon P., André de Palma, and Jacques-François Thisse (1992), *Discrete Choice Theory of Product Differentiation*. Cambridge, MA: MIT Press.

Berry, Steven T. (1994), "Estimating Discrete-Choice Models of Product Differentiation," *RAND Journal of Economics*, 25 (Summer), 242–62.

Bozdogan, Hamparsum (1987), "Model Selection and Akaike's Information Criterion (AIC): The General Theory and Its Analytical Extensions," *Psychometrika*, 52 (September), 345–70.

Bucklin, Randolph E. and Sunil Gupta (1992), "Brand Choice, Purchase Incidence, and Segmentation: An Integrated Modeling Approach," *Journal of Marketing Research*, 29 (May), 201–15.

Chintagunta, Pradeep K. (1992), "Heterogeneity in Nested Logit Models: An Estimation Approach and Empirical Results," *International Journal of Research in Marketing*, 9 (May), 161–75.

——— (1993), "A Product-Market Map From Household Scanner Panel Data: New Brand Introduction and Other Issues," working paper, Cornell University.

———, Dipak C. Jain, and Naufel J. Vilcassim (1991), "Investigating Heterogeneity in Brand Preferences in Logit Models for Panel Data," *Journal of Marketing Research*, 28 (November), 417–28.

Dawes, Robin M. and B. Corrigan (1974), "Linear Models in Decision Making," *Psychological Bulletin*, 81 (2), 95–106.

Fader, Peter S. and James M. Lattin (1993), "Accounting for Heterogeneity and Nonstationarity in a Cross-Sectional Model of Consumer Purchase Behavior," *Marketing Science*, 12 (Summer), 304–17.

Fishbein, Martin (1967), "Attitude and Prediction of Behavior," in *Readings in Attitude Theory and Measurement*, M. Fishbein ed. New York: John Wiley & Sons, 389–99.

Green, Paul E. and Abba M. Krieger (1995), "Attribute Importance Weights Modification in Assessing a Brand's Competitve Potential," *Marketing Science*, 14 (Summer), 253–70.

Grover, Rajiv and William R. Dillon (1985), "A Probabilistic Model for Testing Hypothesized Hierarchical Market Structures," *Marketing Science*, 4 (Fall), 312–35.

Guadagni, Peter M. and John D. C. Little (1983), "A Logit Model of Brand Choice Calibrated on Scanner Data," *Marketing Science*, 2 (Summer), 203–38.

Hardie, Bruce G. S. (1994), "The Market Response to Product Line Extensions," unpublished doctoral dissertation, University of Pennsylvania.

Honnold, Douglas, Robert Brooks, and John D. C. Little (1990), "Logit for the Boardroom," presented at 1990 Marketing Science Conference, Urbana-Champaign, IL.

Johnson, Eric J., Robert J. Meyer, and Sanjoy Ghose (1989), "When Choice Models Fail: Compensatory Models in Negatively Correlated Environments," *Journal of Marketing Research*, 26 (August), 255–70.

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.

——— and ——— (1993), "Measuring Brand Value with Scanner Data," *International Journal of Research in Marketing*, 10 (March), 9–22.

Kannan, P. K. and Gordon P. Wright (1991), "Modeling and Testing Structured Markets: A Nested Logit Approach," *Marketing Science*, 10 (Winter), 58–82.

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

Lancaster, Kelvin (1971), *Consumer Demand: A New Approach*. New York: Columbia University Press.

Lattin, James M. (1990), "Measuring Preference From Scanner Panel Data: Filtering Out the Effects of Price and Promotion," unpublished working paper, Graduate School of Business, Stanford University.

——— and Randolph E. Bucklin (1989), "Reference Effects of Price and Promotion on Brand Choice Behavior," *Journal of Marketing Research*, 26 (August), 299–310.

Lawrence, Richard (1993), "Inside New-Product Statistics: More, or Less ... New, Or Not?" *Journal of Advertising Research*, 33 (March/April), RC3–6.

Little, John D. C. and Eric Anderson (1994), "A Product Choice Model with Marketing, Filtering and Purchase Feedback," unpublished working paper, Sloan School of Management, Massachusetts Institute of Technology.

McFadden, Daniel (1973), "Conditional Logit Analysis of Qualitative Choice," in *Frontiers in Econometrics*, P. Zarembka, ed. New York: Academic Press, 105–42.

Ortmeyer, Gwen, James M. Lattin, and David B. Montgomery (1991) "Individual Differences in Response to Consumer Promotions," *International Journal of Research in Marketing*, 8 (September), 169–86.

Pedrick, James H. and Fred S. Zufryden (1991), "Evaluating the Impact of Advertising Media Plans: A Model of Consumer Purchase Dynamics Using Single-Source Data," *Marketing Science*, 10 (Spring), 111–30.

Quandt, Richard E. (1956), "A Probabilistic Theory of Consumer Behavior," *Quarterly Journal of Economics*, 70 (November), 507–36.

Rangaswamy, Arvind, Raymond R. Burke, and Terence A. Oliva (1993), "Brand Equity and the Extendibility of Brand Names," *International Journal of Research in Marketing*, 10 (March), 61–75.

Russell, Gary J. and Fred M. Feinberg (1995), "The Use of Averaged Purchase History Loyalty Variables in an Inertial Logit Model With Stationary Preferences," paper presented at the 1995 Marketing Science Conference, Sydney, Australia.

Schiller, Zachary (1993), "Procter & Gamble Hits Back," *Business Week*, (July 19), 20–22.

Srinivasan, V. (1979), "Network Models for Estimating Brand-Specific Effects in Multiattribute Marketing Models," *Management Science*, 25 (January), 11–21.