A Segment-level Model of Category Volume and Brand Choice

William R. Dillon • Sunil Gupta
Southern Methodist University
Columbia University

Abstract

It is an everyday marketplace occurrence that brands lose and gain share. However, a brand's sales gain or loss can be attributable to very different factors, and thus understanding the sources of sales gain or loss would seem to be an important aspect of a brand manager's job.

The primary purpose of this research is to develop a model that can answer the following questions: (i) What are the sources of gain or loss of a brand's sales due to category volume and brand switching? (ii) What consumer and marketing characteristics affect consumers' purchase frequency of the product category and that of different brands? (iii) Do all consumers behave similarly, or are there distinct segments which respond to marketing actions differently? (iv) If such segments exist, what is the size and composition of each segment, and what is the appropriate strategy for each group of consumers?

Our modeling approach, which decomposes a brand's sales into category volume and brand choice components, has many similarities and also several differences with traditional approaches. Similar to the NBD and Dirichlet models, we assume that a consumer's category purchase rate follows a Poisson distribution, and the number of purchases per brand follows a multinomial distribution. Our model differs from traditional models by including marketing mix variables, by accounting for loyal or near-loyal consumers, by explicitly incorporating consideration sets, and by segmenting consumers on the basis of their brand perceptions and their responses to marketing mix variables.

We account for consumer hetereogeneity by identifying homogeneous latent segments that capture differences in consumers' response to marketing variables in both brand choice and category volume behavior. Calibration of the category volume and brand choice models and the separation of loyalty and switching segments are done simultaneously so that there is no need to assume any specific hierarchy and, in contrast to the usual assumption of independence between choice and volume which may be unreasonable at the aggregate level, the requirement is that choice and volume decisions be independent only at the segment level.

We investigate the properties of our model in the context of a national survey of supermarket purchases of jumbo paper towels. 2,500 households were surveyed who were asked to provide information on: rolls of paper towels bought in the last 4 weeks, rolls of each brand of paper towel bought in the last 4 weeks, brand usually bought, brands that a household bought or would consider buying, average price paid or expected for each brand bought or considered, ratings on 20 brand attributes (e.g., strength, absorbency, etc.), and demographic information (e.g., family size).

We found that not only did the modeling framework outperform a variety of competing models, it also provided insights into the competitive structure in this market. The loyal segments could be distinguished on the basis of price sensitivity. Less than 10% of the loyal households were price insensitive and, in general, households showed increasing price sensitivity in their category volume decisions if they had more children and were heavier users of paper towels. This was consistent across all brands. The model estimates that about 71% of the households are switchers and five switching segments are needed to characterize household purchase patterns. Two of the largest switching segments (labeled as Price Sensitive and Value Segments) are very price sensitive in both their brand choice and category volume components. In both these segments private label brands are dominant with about 30% share. Segments also emerged on the basis of Strength, Absorbency, and Tearing Ease of paper towels. Interestingly, we found that brand shares within these segments are quite consistent with the objective quality ratings of brands as given by *Consumer Reports*. Finally, price elasticity analysis for one of the brands, Bounty, reveals that a 5% drop in Bounty's price increases its sales by 13.6%. Almost half of this increase comes from brand switching, with the other half coming from increases in category volume (e.g., stockpiling). Bounty gains the most from the price sensitive segments (Price and Value Segments). Private labels are hurt least by Bounty's price cut. (*Buyer Behavior; Choice Models; Promotion*)

Marketing Science/Vol. 15, No. 1, 1996
pp. 38–59

0732-2399/96/1501/0038$01.25

1. Introduction

An article in the *Wall Street Journal* (October 27, 1992) reported that for the 52-week period ending September 27, 1992 sales of Peter Pan peanut butter were $130.5 million, a drop of 11% from the previous year. A few weeks later, the *Wall Street Journal* (November 10, 1992) reported a similar problem facing another brand manager. This time it was Close-Up toothpaste. Its sales in the 52-week period ending September 30, 1992 were $57.7 million, down by 17% compared to the previous year. Clearly, both brand managers have reasons to worry. But are their problems the same? Perhaps not. A 1990–91 drought severely affected the peanut crop, causing peanut butter prices to rise sharply. Consequently sales of the peanut butter category dropped by 8% to $0.91 billion. The toothpaste category, on the other hand, saw a 4% increase to $1.1 billion. Even though the two brands lost sales, the decline in the product category sales perhaps accounted for a large portion of sales loss for Peter Pan, while Close-Up sales loss are perhaps attributable to share erosion. Moreover, these effects are likely to vary by market segments that differ in their brand loyalty and their responsiveness to marketing actions.

The objective of this paper is to develop a model to help brand managers answer the following questions. What are the sources of gain or loss of brand sales (specifically due to category volume and brand switching)? What consumer and marketing characteristics affect consumers' purchase frequency of the product category and that of different brands? Do all consumers behave similarly, or are there distinct segments which respond to marketing actions differently? If such segments exist, what is the size and composition of each segment, and what is the appropriate strategy for each group of customers?

We would like to emphasize that the problems mentioned above and our proposed model are not limited to consumer packaged goods. Similar issues arise in a variety of different contexts. For example, for designing better fare policies and schedules, public transit organizations regularly conduct surveys to find riders' usage frequency for different modes of transportation. Television networks constantly monitor the viewing habits of households by getting information on the number of hours a household spends watching television in, say, a week, and how these hours are spent watching different networks or cable programs. Credit card companies, hotels, airlines, film studios all deal with similar issues of total category usage and usage frequency of different brands or services.

The interest in modeling consumer purchase patterns is not new. The negative binomial distribution (NBD) and the Dirichlet models have been used for this purpose for several years (Morrison and Schmittlein 1988, Goodhardt et al. 1984). Our modeling approach has many similarities and also several differences with these traditional models. For example, similar to the NBD and Dirichlet models, we assume that a consumer's category purchase rate follows a Poisson distribution, and the number of purchases per brand follows a multinomial distribution. Our model differs from the NBD and Dirichlet models by including marketing mix variables, by accounting for loyal or near-loyal consumers, by explicitly incorporating consideration sets, and by segmenting consumers based on their brand perceptions and their response to marketing mix variables. Unlike NBD and Dirichlet models, we account for consumer heterogeneity by identifying homogeneous latent segments that capture differences in consumers' response to marketing variables in both brand choice and category volume behavior. This approach offers three distinct advantages. First, we do not have to make any distributional assumptions about consumer heterogeneity. Second, consumers are allowed to have different sensitivities to marketing variables. Third, the usual assumption of independence between choice and volume, which may be unreasonable at the aggregate level, is required only at the segment level.

In most product categories, many consumers are very loyal to a particular brand. Many researchers have addressed this issue by creating a brand loyal or nearly-loyal segment in an ad-hoc fashion. For example, consumers may be defined as loyal to a brand if in a given time period 90% or more of their purchases are of that brand. Defining consumers as brand loyal based on their observed purchase behavior may be erroneous. For example, if we toss a fair coin 3 times and observe 3 heads in a row, it does not imply that the probability of head in a toss is one. Consumers who appear loyal may actually be switchers, and it may simply be by chance that we observe them buying the same brand n

times (Colombo and Morrison 1989, Grover and Srinivasan 1992). Clearly, as n increases we are more and more certain that this consumer is really loyal to that brand. Since the loyal consumers create a spike close to choice probability of 1, they need special consideration—much like the consumers who create a spike at zero in the NBD models (Morrison and Schmittlein 1988). Some studies have handled loyal consumers by excluding them from further analysis. For example, Kamakura and Russell (1989) excluded 31.5% of their sample on this basis. If our objective is to simultaneously include brand choice and category consumption behavior, exclusion of these loyal consumers can potentially bias our results for the category purchase model (e.g., loyal customers may also be heavy users of the category).

Colombo and Morrison (1989) proposed a simple and elegant model that uses aggregate switching matrix to assess the proportion of hard-core loyals and potential switchers. We build on Colombo and Morrison approach in several ways. We include marketing mix variables. We also propose that all switchers are not alike. For example, some may be more price sensitive than others. Therefore we allow further segmentation of switchers. We also include category volume decisions of consumers. Therefore consumer segments are derived based on both the choice and volume behavior of consumers.

Krishnamurthi and Raj (1988) and Tellis (1988) suggested a model of brand choice and brand quantity. In a follow up study, Krishnamurthi and Raj (1991) further enhanced their approach by a priori segmenting consumers into loyal and non-loyal groups. While a priori segmentation may be appropriate in some cases, recent studies have suggested that it is generally more desirable to uncover segments based on consumers' response to marketing actions (Kamakura and Russell 1989, Bucklin and Gupta 1992). Further, Krishnamurthi and Raj (1991) approach does not account for consumers' response heterogeneity in their quantity decision. Finally, as Grover and Srinivasan (1992) suggest, a model of brand quantity only captures the stockpiling effect due to price and promotion, but ignores the effect on consumer purchase timing. They argue that a category volume model captures both the stockpiling and purchase timing effects simultaneously. In other words, while Krishnamurthi and Raj (1991) capture only the stockpiling effect, and Bucklin and Gupta (1992) capture only the purchase time effect, the proposed model with category volume captures both these effects.

Grover and Srinivasan (1992) evaluated the effect of retail promotions on brand share and category sales. They propose a market segmented as brand loyal and brand switchers. Our study differs from Grover and Srinivasan (G&S) in three important ways. First, G&S group consumers as brand loyal or switchers based on their observed brand shares. Note that brand choice of a consumer is affected by his intrinsic preference *and* his sensitivity to marketing mix elements such as price. Therefore two consumers can have similar observed brand shares even if they have different brand preferences and price sensitivities. From a managerial perspective it is perhaps more relevant to group consumers based on their sensitivity to marketing actions. Our study segments consumers based on this objective. Second, unlike G&S we further segment brand loyal and brand switchers based on their category purchase behavior. In other words, we do *not* assume that category price and promotion will have identical impact on the category purchase volume of all consumers who belong to a brand loyal or brand switching segment. In fact, our empirical applications show that this assumption is unlikely to hold. Third, unlike G&S we simultaneously estimate segment sizes, and brand choice and category volume models. A sequential approach as adopted by G&S (choice followed by category sales), which includes inclusive or category attractiveness value, is generally inefficient (Ben-Akiva and Lerman 1985).

The paper is organized as follows. We begin with a description of the model in Section 2. The model is discussed in terms of its two components, the category volume model and the brand choice model. Estimation issues are discussed in Section 3. Section 4 presents an application of our model to a recent data set for the paper towels category. To provide a proper assessment of the model, we also discuss the results from six nested models and five competing models. We discuss the managerial implications of these results in Section 5 and end with conclusions in Section 6.

2. Model

We begin with a brief description of the consumer purchase behavior and our modeling approach. Next, we describe individual-level category volume and brand choice models, which are followed by a discussion of the segment-level models.

2.1. Consumer Purchase Behavior

A consumer's purchase behavior is determined by his/her need for the product category (e.g., a larger family may have a higher category purchase rate), his/her brand preference (e.g., a consumer is likely to buy more of his favorite brand), and the marketing conditions (e.g., the price and promotion conditions in the market place are likely to influence both the brand choice and the category volume decisions of a consumer). In general, a consumer's category volume and brand choice decisions are likely to be influenced by different factors. For example, family size may affect a consumer's category volume decision but may have no influence on his brand choice decision. Therefore it is useful, both from the perspective of consumer purchase behavior and from the perspective of managers (as discussed earlier), to decompose the brand sales into category volume and brand choice components.

Our approach allows a consumer to buy more of his favorite brand in two ways. First, for a given category volume, the higher the choice probability for a certain brand, the greater the expected purchase quantity for that brand. Second, we let the marketing actions of a consumer's favorite brand affect his category volume decision more than the marketing activity of his less favored brands by defining category level variables (e.g., price) as the weighted sum of brand prices, where the weights reflect a consumer's brand preference. This allows the price of a consumer's favorite brand to affect his category volume, and hence the volume of the favorite brand, more than the price of other brands.

Researchers have assumed different ordering of consumer's choice process. For example, Krishnamurthi and Raj (1988) model is consistent with the assumption that consumers make brand choice decision first, followed by their decision of brand quantity. Several other models (e.g., NBD-Dirichlet models) assume that choice and volume decisions are independent. This implies that brand shares do not vary across different usage groups. This is clearly suspect at the aggregate level. In contrast, our segment-level approach assumes independence of these two components within a segment. When aggregated, the two components can, and generally do, exhibit dependencies.

2.2. The Modeling Approach

Let $\mathbf{y}_h = (y_{h1}, y_{h2}, \ldots, y_{hJ})$ denote the vector of purchase frequencies of J brands for consumer h in a given time interval. Further $y_h = \Sigma_{j=1}^{J} y_{hj}$ gives the total purchase frequency for the product category for consumer h. Many data collection instruments (e.g., surveys) use screening questions to decide which respondents "qualify" to be included in the sample. Typically, respondents who have not bought the product in a specified time interval are excluded.[1] Therefore the probability that consumer h buys $(n_{h1}, \ldots, n_{hJ})$ units of brands 1, 2, $\ldots, J$ given non-zero purchases of the category can be represented by $\Pr(y_{h1} = n_{h1}, \ldots, y_{hJ} = n_{hJ} | y_h > 0)$. This probability can be broken into two components as follows.

$$\begin{aligned} \Pr(y_{h1} &= n_{h1}, \ldots, y_{hJ} = n_{hJ} | y_h > 0) \\ &= \Pr(y_h = n_h | y_h > 0) \\ &\quad \times \Pr(y_{h1} = n_{h1}, \ldots, y_{hJ} = n_{hJ} | y_h = n_h) \qquad (1) \end{aligned}$$

where $n_h = \Sigma_j n_{hj}$. The first component captures the category purchase behavior of the consumer (*category volume model*), and the second component describes his/her allocation of total category purchases across different brands (*brand choice model*).

2.3. Category Volume Model

The first component of equation (1) describes the probability that consumer h purchases the product category n_h times in the pre-specified time interval, given that he buys at least once. Many researchers have suggested that consumers' category purchase behavior can be captured by a Poisson process, which in spite of its simplicity, "remains a workable approximation," (Goodhardt, et al. 1984, p. 626). Therefore category purchases

[1] These "zero purchasers" are excluded from the sample because they are usually infrequent category buyers and hence not important to most companies. This method may exclude some heavy buyers who made zero purchases (simply by chance) in the specified time period. However, the higher the purchase rate of a consumer, the lower is the probability of his making zero purchases in a given time interval. Note, if zero purchases are included in the sample, the model can be easily modified.

of consumer h with purchase rate λ_h can be represented as

$$\Pr_{\mathcal{P}}(y_h;\lambda_h) = \Pr(y_h = n_h \mid y_h > 0) = \left(\frac{\lambda_h^{n_h} e^{-\lambda_h}}{n_h!}\right)\left(\frac{1}{1-e^{-\lambda_h}}\right) = \frac{\lambda_h^{n_h}}{(e^{\lambda_h}-1)n_h!}. \quad (2)$$

Further, to account for the influence of consumer characteristics and marketing mix variables, the purchase rate parameter is expressed as

$$\lambda_h = \exp(\mathbf{z}_h'\boldsymbol{\gamma}) \quad (3)$$

where $\mathbf{z}_h$ is the vector of covariates and $\boldsymbol{\gamma}$ is the vector of parameters. We use an exponential function on the right hand side of equation (3) to ensure that the purchase rate is strictly positive.

2.4. Brand Choice Model

If each consumer is assumed to follow a zero order process and the number of purchases in the product category is given, then the brand purchase frequencies follow a multinomial distribution. The assumption of zero order has been used extensively in the marketing literature because of its mathematical tractability and empirical support (Morrison and Schmittlein 1988). Therefore the second component of equation (1) can be written as:

$$\Pr_{\mathcal{M}}(\mathbf{y}_h; n_h, \mathbf{p}_h) = \Pr(y_{h1} = n_{h1}, \ldots, y_{hJ} = n_{hJ} \mid y_h = n_h) = \frac{n_h!}{\prod_j n_{hj}!}\prod_j p_{hj}^{n_{hj}} \quad (4)$$

where p_{hj} is the probability that consumer h buys brand j.

Following Colombo and Morrison (1989), we assume that heterogeneity in consumer choice can be captured by loyal and switching segments. If θ_j is the relative size or share of brand j's loyal segment (i.e. θ_j represents the likelihood of finding a consumer in brand j's loyal segment), θ_k is the share of brand k's loyal segment, and θ_s is the share of the switching segment, then

$$\begin{aligned} p_{hj} &= \Pr(j \mid \text{consumer } h \text{ is loyal to brand } j)\cdot\theta_j \\ &\quad + \sum_{k\neq j}\Pr(j \mid \text{consumer } h \text{ is loyal to brand } k)\cdot\theta_k \\ &\quad + \Pr(j \mid \text{consumer } h \text{ is a switcher})\cdot\theta_s \\ &= 1\cdot\theta_j + \sum_{k\neq j} 0\cdot\theta_k + p_{hj|s}\cdot\theta_s = \theta_j + \theta_s\cdot p_{hj|s} \quad (5) \end{aligned}$$

where $p_{hj|s}$ is the probability that consumer h buys brand j given that he is a switcher (in our subsequent discussion, we further segment the switchers). Notice that $0 < \theta_j, \theta_s \leq 1$, and $\Sigma_j\, \theta_j + \theta_s = 1$. If a consumer is not completely loyal to any one brand, then the choice probability $p_{hj|s}$ can be captured by a logit model (Gupta 1988):

$$p_{hj|s} = \frac{\exp(\mathbf{x}_{hj}\boldsymbol{\beta}_s)}{\Sigma_k \exp(\mathbf{x}_{hk}\boldsymbol{\beta}_s)} \quad (6)$$

where $\mathbf{x}_{hj}$ is the vector of covariates and $\boldsymbol{\beta}_s$ is the vector of response parameters for the switching segment.

2.5. Segment-level Model

It is important to recognize heterogeneity in consumers' purchase behavior. We have already alluded to this heterogeneity in our brand choice model by introducing loyal and switching segments. However, not all switchers are alike—for example, some may be more price sensitive than others. Further, we need to account for differences in consumers category purchase rates as well. Typically these heterogeneities are incorporated by assuming that choice probabilities and purchase rates are distributed across consumers according to a prespecified distribution (e.g., Dirichlet for choice probabilities, and Gamma for purchase rate, see Goodhardt, et al. 1984). An alternative approach, which we use, is to group consumers based on how they respond to marketing activities such as price. This latent class approach allows for a semi-parametric distribution of heterogeneity that is far more flexible than any pre-specified distribution. This response-based segmentation is managerially more useful than a demographic segmentation (Bucklin and Gupta 1992).

Conceptually our segmentation scheme can be visualized as follows. If there are J brands in the market, then consumers are grouped into J loyal segments and one switching segment. Each loyal segment is further divided into G_j segments based on category purchase behavior of consumers. The switching segment is also divided into G_S segments based on both brand choice and category purchase behavior. In other words, one can visualize all switchers being split into multiple switching segments based on their response parameters in the choice model. Each of these switching segments is further broken into multiple segments based on the response parameters in the category purchase model. Notice that this hierarchical description of segmentation (i.e. choice first, then category) is only for illustrative purposes. The actual estimation, as described below, is done simultaneously on both choice and category purchase. Therefore no specific hierarchy is assumed.

The likelihood function for a household, $\mathcal{L}(\mathbf{y}_h)$, can therefore be written as the weighted sum of segment-level likelihoods, where the weights are the unknown segment shares.

$$\mathcal{L}(\mathbf{y}_h) = \sum_j \theta_j \delta_{hj} \sum_{g_j=1}^{G_j} w_{g_j} \frac{\lambda_{h|g_j}^{n_h}}{(e^{\lambda_{h|g_j}} - 1)n_h!} + \theta_s \sum_{g_s=1}^{G_s} w_{g_s} \left(\frac{n_h!}{\prod_j n_{hj}!} \prod_j p_{hj|g_s}^{n_{hj}} \right) \frac{\lambda_{h|g_s}^{n_h}}{(e^{\lambda_{h|g_s}} - 1)n_h!}$$
$$= \mathcal{L}_{\text{loyal}} + \mathcal{L}_{\text{switcher}} \qquad (7)$$

where $\Sigma_j \theta_j + \theta_s = 1$, $\Sigma_{g_j} w_{g_j} = 1$, $\Sigma_{g_s} w_{g_s} = 1$, $\lambda_{h|g_j}$ and $\lambda_{h|g_s}$ are purchase rates for household h given that it belongs to loyal segment g_j, or to switching segment g_s, and δ_{hj} is 1 if all purchases of household h are of brand j, otherwise it is 0. The overall likelihood can be broken into two components: one relating to loyal segments ($\mathcal{L}_{\text{loyal}}$) and the other relating to the switching segments ($\mathcal{L}_{\text{switcher}}$). If all purchases of a consumer are of brand j ($\delta_{hj} = 1$), he may be loyal to j or a switcher who bought all brand j by chance. However, if even one of the purchases of this consumer are of any other brand k, then by our definition, he can not be loyal to j. In such a case, $\mathcal{L}(\mathbf{y}_h) = \mathcal{L}_{\text{switcher}}$.

The likelihood for loyal segments, $\mathcal{L}_{\text{loyal}}$, consists of the indicator variable δ_{hj}, and the likelihood for category purchase segment. These two components are weighted by the respective segment sizes θ_j and w_{g_j}. Here θ_j is the share of brand j's loyal segment, and w_{g_j} represents the share of category purchase segments within brand j's loyal segment.

In general, θ_j and w_{g_j} are estimated simultaneously unless a priori information on these parameters is available. When a simultaneous estimation is used, θ_j and w_{g_j} are not separately identifiable. Only the product $\pi_{g_j} = \theta_j w_{g_j}$ is estimable. Letting $l = 1, 2, \ldots, L$ denote the total number of loyal segments ($L = \Sigma_j G_j$) and $s = 1, 2, \ldots, S$ denote the total number of switching segments ($S = \Sigma_s G_s$) we can rewrite equation (7) as

$$\mathcal{L}(\mathbf{y}_h) = \sum_{l=1}^{L} \pi_l \delta_{hj} \left(\frac{\lambda_{h|l}^{n_h}}{(e^{\lambda_{h|l}} - 1)n_h!} \right) + \sum_{s=1}^{S} \pi_s \left(\frac{n_h!}{\prod_j n_{hj}!} \prod_j p_{hj|s}^{n_{hj}} \right) \left(\frac{\lambda_{h|s}^{n_h}}{(e^{\lambda_{h|s}} - 1)n_h!} \right) \qquad (8)$$

where $p_{hj|s}$ is defined in (6), and

$$\lambda_{h|l} = \exp(\mathbf{z}_h' \boldsymbol{\gamma}_l), \quad \text{and} \quad \lambda_{h|s} = \exp(\mathbf{z}_h' \boldsymbol{\gamma}_s). \qquad (9)$$

The model described above can be viewed as an extension of the basic latent class model (Lazarsfeld and Henry 1968) with two important differences. First, the proposed model groups consumers not only on the basis of their choice probabilities but also on the basis of their purchase rates. More information on each consumer is used for the purpose of assigning consumers to their respective latent segment. This additional information allows estimation of certain models that otherwise would be unidentified. Second, the proposed model distinguishes between loyal and switching consumers.

3. Estimation

3.1. Finite Mixture Model

In general, the unconditional likelihood for a household with purchase vector $\mathbf{y}_h$ can be written as

$$\mathcal{L}(\mathbf{y}_h) = \int \mathcal{L}(\mathbf{y}_h | \varphi) dF(\varphi) \qquad (10)$$

where $\mathcal{L}(\mathbf{y}_h | \varphi)$ is the conditional likelihood with parameters φ, and $F(\cdot)$ is the mixing distribution. It can be shown that a continuous mixing distribution function $F(\cdot)$ can be consistently estimated with a finite number of K mass points (cf. Simar 1976), i.e.

$$\mathcal{L}(\mathbf{y}_h) = \sum_{k=1}^{K} \alpha_k \mathcal{L}(\mathbf{y}_h | \varphi_k) \qquad (11)$$

where φ_k is the vector of parameters, and α_k is the mixing weight or segment share for segment k, such that $0 < \alpha_k \le 1$ and $\Sigma_k \alpha_k = 1$. Equation (11) is a simplified representation of equation (8) where α_k is equivalent to $\{\pi_l \text{ and } \pi_s\}$, and φ_k is equivalent to $\{\boldsymbol{\gamma}_l, \boldsymbol{\gamma}_s, \text{and } \boldsymbol{\beta}_s\}$. The accumulating empirical and theoretical evidence strongly suggests that the number of mass points required to characterize the mixing distribution is often very small (Card and Sullivan 1988). Many algorithms can be used to find the maximum likelihood solutions. We use the EM algorithm described below. To facilitate discussion, we will use the simplified notation of (11) in this section.

3.2. EM Algorithm

The EM algorithm uses the following two steps (Dempster, Laird and Rubin 1977).

The E-step. In the E-step (expectation step) new estimates of the segment membership probabilities (i.e., the posterior probabilities) are obtained based upon provisional estimates of α_k and φ_k. Specifically,

$$\alpha_{k|\mathbf{y}_h} = \frac{\alpha_k \mathcal{L}(\mathbf{y}_h | \varphi_k)}{\mathcal{L}(\mathbf{y}_h)} \tag{12}$$

The M-step. In the M-step (maximization step), the "complete" data log-likelihood function is maximized to obtain new estimates of α_k and φ_k. The "complete" data log-likelihood function is captured by introducing indicator random variables I_{hk} which is 1 if $h \in k$, 0 otherwise. Further, assume that I_{hk} are multinomially distributed such that

$$\Pr(I_{h1}, \ldots, I_{hK} | \alpha_1, \ldots, \alpha_K) = \prod_{k=1}^{K} \alpha_k^{I_{hk}} \tag{13}$$

Once the I_{hk} are introduced the conditional likelihood for a household can be written as

$$\mathcal{L}(\mathbf{y}_h | I_{h1}, \ldots, I_{hK}) = \prod_{k=1}^{K} I_{hk} \Pr(\mathbf{y}_h | \varphi_k)$$

$$= \prod_{k=1}^{K} \Pr(\mathbf{y}_h | \varphi_k)^{I_{hk}} \tag{14}$$

Using (13) and (14), we can write the unconditional likelihood for a household as

$$\mathcal{L}(\mathbf{y}_h) = \prod_{k} \Pr(\mathbf{y}_h | \varphi_k)^{I_{hk}} \alpha_k^{I_{hk}} \tag{15}$$

Therefore, the "complete" data log-likelihood across all households is

$$\mathcal{LL} = \sum_{h=1}^{H} \sum_{k=1}^{K} I_{hk} \{\ln \Pr(\mathbf{y}_h | \varphi_k) + \ln \alpha_k\}$$

$$= \sum_{h=1}^{H} \sum_{k=1}^{K} I_{hk} \ln \alpha_k + \sum_{h=1}^{H} \sum_{k=1}^{K} I_{hk} \{\ln \Pr_{\mathcal{P}}(y_h; \lambda_{h|k})$$

$$+ \ln \Pr_{\mathcal{M}}(\mathbf{y}_h; n_h, \mathbf{p}_{h|k})\} \tag{16}$$

where $\Pr(y_h | \varphi_k)$ is the product of the category volume Poisson probability, and brand choice multinomial probability for segment k, and $\lambda_{h|k}$ and $\mathbf{p}_{h|k}$ are parameterized in terms of covariates. Before recycling back to the E-step, an estimate of α_k is obtained by

$$\hat{\alpha}_k = \frac{1}{H} \sum_{h=1}^{H} \hat{\alpha}_{k|\mathbf{y}_h} \tag{17}$$

The parameter estimates from the M-step are used in subsequent E-steps to arrive at updated estimates of the unobserved data (i.e., the I_{hk}'s). Armed with the new estimates of the unobserved data, the subsequent M-step computes updated estimates of all within-segment parameters and segment probabilities. This process, alternating E-step and M-step, is continued until the improvement in log likelihood is less than some convergence criterion.

It is important to note that the log-likelihood function shown in equation (16) is expressed as an additive function of the Poisson and multinomial components. The additivity of these two components has been recently proved by Dzhafarov and Böckenholt (1993). The important implication of this result on estimation is that the parameters φ_k and α_k's can be estimated separately, which acutely reduces the computational burden and results in more stable parameter estimates, especially in the case of large-scale problems. For example, our proposed model estimates 151 parameters. However, due to the separability property, the estimation procedure can divide the total number of parameters into 22 segment size parameters, 69 volume-related parameters, and 60 choice-related parameters.

Because the normal equations associated with the parameters are non-linear, numerical optimization methods must be used within each M-step. Several well-known options are available. The program we used allows the user to choose among three estimation methods: Newton Raphson, BHHH (Berndt et al. 1974), and a modified simplex method. The Newton-Raphson method requires first and second order derivatives of the log-likelihood function (see Appendix 1 for derivatives for our model). The BHHH method requires only (outer products of) the first order derivatives of the log-likelihood function. The modified simplex method is an adaptation of the algorithm suggested by Simar (1976), and recently used by Brännäs and Rosenqvist (1994) in the context of heterogeneous count data. The simplex method, implemented with the (modified) NSIMP subroutine available in GQOPT, does not require gradient information, and can be used in several different ways. If it is used to provide (final) parameter estimates then after convergence is obtained, standard errors of the parameters estimates can be obtained by use of the BHHH-algorithm.

The general form of the Poisson-multinomial model is identified (see Böchenholt 1993a, and Dzhafarov and Böchenholt 1993), but not all variations may be. Because the EM algorithm guarantees convergence to a local maximum, our program reports information on the rank of the covariance matrix of parameter estimates. Our experience with estimating this class of models suggests that after the first M-step, which generally involves a relatively large number of iterations, convergence is usually obtained after a few (very often less than 10) iterations within each successive M-step since parameter estimates from the previous iteration are used as starting values. However, a large number of EM cycles is generally needed.

3.3. Choosing the Number of Segments

The estimation procedure requires that the number of segments or latent classes K be specified a priori. However, in most applications, K is unknown and therefore needs to be inferred from the data. In finite mixture models, the regularity conditions necessary for the traditional likelihood ratio test for K do not hold (Titterington, et al. 1985). Therefore, we use the following two methods to determine K.

In the first method we select K, which minimizes the CAIC_K criterion (Bozdogan 1987):

$$\text{CAIC}_K = -2\mathcal{LL}_K + Q_K[\ln H + 1] \qquad (18)$$

where H is the number of households, $\mathcal{LL}_K$ is the log-likelihood, and Q_K is the number of parameters for K segments. As K increases, the $\mathcal{LL}$ generally improves, but it also requires estimating more parameters. Similar to adjusted R^2, CAIC strikes a balance between these two opposing factors.

In the second method, the criterion for a global maximum of the overall likelihood is that the function

$$A(\varphi) = \sum_{h=1}^{H} \frac{\mathcal{L}(\mathbf{y}_h \mid \varphi)}{\mathcal{L}(\mathbf{y}_h)} - H \leq 0. \qquad (19)$$

for all values $\varphi \in \Omega$. The maximizing estimates φ^* have the property that $A(\varphi^*) = 0$ and $A(\varphi) \leq 0$, $\forall \varphi \in \Omega$ (Lindsay 1983, Brännäs and Rosenqvist 1992).

A simple example will clarify the intuition behind equation (19). Consider a market of 100 consumers, where each consumer makes 2 purchases in a given time period. Further, assume that this market consists of two segments where 50 consumers have a probability 1 of buying a brand, and the other 50 have a probability 0. To estimate the brand choice probability (φ) using observed purchases, we start with $K = 1$ and get a maximum likelihood estimate of $\varphi^* = 0.5$. The denominator of equation (19), $\mathcal{L}(\mathbf{y}_h)$, is then evaluated for each household h at $\varphi^* = 0.5$. The numerator, $\mathcal{L}(\mathbf{y}_h \mid \varphi)$ is evaluated for each household at different values of φ in the parameter space (in this case 0-1). It is easy to see that in this example $A(\varphi) > 0$ for some values of φ, e.g.,

$$A(\varphi = 1.0) = 50\left[\frac{(1)^2}{(0.5)^2}\right] + 50\left[\frac{(0)^2}{(0.5)^2}\right] - 100$$
$$= 100 > 0, \qquad (20)$$

which suggests that we need more than one segment. The specific steps of this method are as follows.

1. Set $K = 1$ and $\alpha_1 = 1$ and maximize $\mathcal{LL}$ with respect to all unknown parameters, over a reasonable (but relatively wide) range of $\hat{\varphi}$.

2. For fixed parameter estimates, scan all $A(\hat{\varphi})$ to determine if the condition $A(\hat{\varphi}) \leq 0$ holds. If it does, then stop the estimation procedure—this yields the maximum likelihood estimator.

3. If $A(\hat{\varphi}) > 0$, increase K by one and add a support point φ_K at the maximum of $A(\hat{\varphi})$—that is, locate the interval in Ω where $A(\hat{\varphi}) > 0$ and add a support point to $\hat{\varphi}$ over those intervals. Re-estimate all the unknown parameters and return to step 2.

4. Application

Our application is based on a recent survey conducted for the paper towels category. We illustrate our model with survey data instead of scanner panel data for several reasons. First, as indicated in the introduction, our model is designed for a general problem that is not limited to consumer packaged goods. For many non-packaged goods, such as consumer durables, industrial products, services, and pharmaceutical, scanner panel data are not available. On the other hand, surveys are still heavily used in the packaged goods industry. Second, even for packaged goods, household scanner panels have limitations—they are available in only a few markets and do not account for product sales through convenience stores and warehouse stores. Third, scanner panels generally do not collect perceptual and attitudinal information from consumers. As shown in our empirical application, attribute perception information can be very useful in providing a better understanding of the market. Fourth, panels typically suffer from attrition of panel members, which, if not properly adjusted for, can lead to severe biases in parameter estimates (Winer 1983). Fifth, Heckman and Willis (1977, p. 32) note that if the model assumes the errors over time to be independent (which is typically the case in most scanner studies), then panel data are unnecessary. In such cases, successive cross-sections, or even a single cross section, is sufficient to estimate the model appropriately. Finally, studies in several fields, such as transportation, extensively use surveys to estimate demand models.

Clearly, surveys have their limitations. For example, the lack of longitudinal data precludes the modeling of any dynamics. However, note that dynamic effects come from three main sources: (a) changes in parameters over time—since this is a long term issue which requires several years of data, almost all studies in marketing do not address this issue; (b) error correlation—again most scanner studies ignore this, which in fact can lead to incorrect estimates as shown by time series studies in econometrics; and (c) lagged effect, such as the effect of inventory on future category purchases. It is this last effect that is ignored in cross-sectional surveys. It is possible to capture this effect by multiple waves of surveys. More importantly, most studies using disaggregate household level data have found the effect of inventory on purchase quantity to be significant but extremely small (e.g., Gupta 1988). Another limitation of surveys is the measurement error due to self-reported nature of the data. Three points should be noted here. First, some previous marketing studies (e.g., Krishnamurthi and Raj 1988) have used diary panel data that suffer from similar measurement problems. Second, recent studies suggest that scanner systems have a significant error in reporting prices (e.g., Goodstein 1994). Some estimates suggest that store scanner data (which provides causal information for household panels) have approximately 75,000 errors per week (Kruger 1990). Although these errors are scrutinized through sophisticated computer systems, many errors go undetected. Third, even if the scanner data measure the marketing environment (e.g., price) more objectively, from a consumer behavior perspective it is the perceived price and not the objective price that really affects consumer decisions. We should note that in our application the average reported prices and share of different brands match closely with the prices and shares of these brands as reported by IRI store scanner data. The average share difference between our survey and IRI scanner data was 0.54%, with a range of 0.2%–0.9% share points. The average price difference between the two data sources was 4 cents (or 5% of the price), with a range of 2–10 cents (except for one brand with a 10 cent price difference, *all* other brands had price difference of less than 5 cents).

In sum, both surveys and scanner data have their advantages and disadvantages. However, given the general nature of the problem being addressed here and a wider application of surveys across different industries, products and services, we chose to illustrate our model with survey data. We start with a brief description of the data, followed by variables, model comparison, and results.

4.1. Data

The data for this application are based on a national survey conducted by a marketing research company in 1991. The focus of this survey was on the supermarket purchases of jumbo paper towels, bought in the last 4 weeks. The survey was restricted to jumbo paper towels since this size accounts for the majority of sales in the paper towels category.[2] 2500 households were surveyed who were asked the following questions: rolls of paper towels bought in the last 4 weeks, rolls of each brand of paper towels bought in the last 4 weeks, brand usually bought, brands that a household bought or would consider buying, average price paid or expected for each brand bought or considered, ratings on 20 brand attri-

[2] According to a report from IRI, of the total paper towels product category, jumbo paper towels accounted for 84.4% of unit sales in the 52 weeks ending 9/6/92.

butes (e.g., strength, absorbency etc.), and demographic information (e.g., family size).

Respondents mentioned a total of 13 brands in the survey. However the top 7 brands accounted for about 76% of the total volume purchased. We restricted our analysis to these 7 brands. Consequently, households who did not buy one of these 7 brands as their usual brand were eliminated. Our final sample was 2177 households instead of the 2500 households originally surveyed. In this sample Bounty was the market leader followed by a strong presence of the private label or store brands. The average purchase rate for the sample was 4.1 rolls of paper towels per month. Households had an average consideration set size of 3.85 brands, with the mode and median at 4 brands. Descriptive statistics for the data are given in Table 1.

4.2. Variables

Category volume. The following variables are used in the category volume model.

Price: Price for the category volume model is the weighted average price of brands where weights are household level shares of brands in the consideration set of that household. We expect category price to have a negative coefficient.

Number of Children: Industry experts believe that paper towels consumption increases significantly with the number of children in the household. Further, economic theory suggests that heavy users (in this case households with many children) are likely to be more price sensitive than light users. If these expectations hold in our analysis they will provide a good targeting strategy, especially for economy brands such as private labels.

Brand choice. The following variables are used in the brand choice model.

Brand Attributes: Respondents rated all the brands they bought or would consider buying on 20 attributes. These attributes were factor analyzed and 5 factors were retained. These factors are strength, absorbency, tearing ease, all-purposeness, and economy. Different brands stress different attributes. For example, Bounty is positioned as a brand with strength and absorbency, whereas private labels are typically viewed as economy brands. If segments emerge based on perceptual attributes, it will be interesting to see if brand shares within these segments reflect such positioning strategies of brands.

Price: This is the price paid or expected for each brand in the consideration set of a household. We expect price

Table 1 Descriptive Statistics for Paper Towel Data

	Households[1]		Volume[2]		Loyal hh[3]			
Brands	No.	%	No.	%	No.	Volume	Consideration Mentions[4]	Average Price ($)
Private Label	417	19.15	2011	22.53	200	628	1905	0.75
Bounty	547	25.13	2132	23.38	259	844	1655	0.87
Brawny	316	14.52	1247	13.97	126	442	1308	0.82
Scott	344	15.80	1295	14.51	120	315	1310	0.84
Viva	231	10.61	894	10.01	69	270	1465	0.82
Hi-Dri	201	9.23	826	9.25	76	231	855	0.77
Mardi Gras	121	5.56	522	5.85	42	137	545	0.80
Total	2177	100.00	8927	100.00	892	2867		0.82

[1] These columns represent the households in our sample who *usually* buy a given brand. For example, of the total 2177 households surveyed, 417 households indicated that they usually buy a private label.

[2] Volume represents the rolls of paper towel of a particular brand bought in the last 30 days. For example, the entire sample of 2177 households reported buying 2011 rolls of private label brand in the last 30 days.

[3] These columns represent the number and purchase volume of households who *appear* loyal, i.e. they bought the same brand in the last 30 days. For example, 200 households reported buying only private label brand in the last 30 days. These households bought 628 rolls of paper towel.

[4] This column shows the number of households who stated that they would consider buying a given brand. For example, of the total 2177 households surveyed, 1905 households stated that they have bought or would consider buying a private label paper towel.

to have a significant impact on consumers' brand choice.

In addition to the variables mentioned above, data were collected on consumers' gender, age, and occupation. However, there is no reason to believe that consumers category volume or brand choice behavior differ by these demographic variables.

4.3. Proposed Model: Number of Segments and Goodness of Fit

4.3.1. Number of segments. The proposed model was estimated using the procedure described in section 3. Based on this method we selected 3 segments for the households loyal to Private Labels, Bounty, Brawny, and Viva; 2 segments for the households loyal to Scott, Hi-Dri, and Mardi Gras; and 5 segments for the switching households. We obtained the same results using the CAIC and the $A(\varphi)$ criteria.

4.3.2. Goodness of fit, stability, and predictive validity. We evaluate the goodness of fit of the model based on log-likelihood, CAIC, and the aggregate and disaggregate total absolute error (TAE) which are defined as

$$\text{Aggregate TAE} = \sum_{j=1}^{J} \left| \sum_{h=1}^{H} \hat{y}_{hj} - \sum_{h=1}^{H} y_{hj} \right| \qquad (21)$$

$$\text{Disaggregate TAE} = \sum_{h=1}^{H} \sum_{j=1}^{J} |\hat{y}_{hj} - y_{hj}| \qquad (22)$$

where y_{hj} is the actual quantity of brand j bought by household h, and $\hat{y}_{hj}$ is the model predicted quantity. Disaggregate TAE determines how good the model is in estimating purchase quantity of each brand for *each household*, while aggregate TAE evaluates model fit for each brand *across all households*. Prediction ability of the model is also evaluated using aggregate and disaggregate TAE for a holdout sample of households.

The model fits the data very well. The aggregate and disaggregate total absolute error (TAE) were 135.2 and 446.3 respectively, which, based on the total category volume of 8927 rolls, represent a percentage error of 1.5% (aggregate) and 5.0% (disaggregate).

To assess the stability of the parameters, the sample was randomly split in half and the model was estimated for each half. In each of the half-samples, the parameter estimates obtained were extremely close to the estimates obtained from the entire sample. The correlation between respective pairs of parameter estimates ranged from a low of 0.81 to a high of 0.94. The parameter estimates obtained in each of the half-samples were also consonant. The correlation ranged from a low of 0.79 to a high of 0.91. A test of the hypothesis of no differences between all respective pairs of parameter estimates could not be rejected at the 0.05 significance level.

To assess the predictive validity of the model, we randomly selected 550 respondents (about 25% of the sample) as a holdout sample who purchased a total of 2250 rolls of paper towels. The model was recalibrated for the remaining sample of 1627 respondents. The parameter estimates obtained were then used to predict the brand volumes of the respondents in the holdout sample. The results revealed aggregate TAE of 78.7 or 3.5%, and disaggregate TAE of 194.6 or 8.6%.

4.4. Model Comparison

In order to evaluate the usefulness of the proposed model, it is desirable to compare it with a few nested and competing models. In this section we propose a series of nested models that highlight the incremental value of each model component. We also describe five competing models against which our model is compared.

4.4.1. Nested models. The nested models are constrained versions of the proposed model and, hence, need fewer parameters. Table 2 gives the likelihood function and Table 3 provides the goodness of fit and prediction results for these nested models.

Nested Model 1 (N_1) includes a reduced set of choice-related covariates. Specifically, we exclude five perceptual attributes from the full model.[3] A likelihood ratio test of N_1 versus the proposed model strongly rejects N_1 ($\chi^2 = 1488$ with 25 degrees of freedom). Other statistics, such as CAIC and TAE also support this result. For example, based on aggregate TAE for the total sample, model N_1 is $(162.9 - 135.2)/135.2$ or 20% worse than

[3] It is possible that the nested models have different number of segments than the proposed model. Therefore, for each nested model we started with the segmentation scheme of the proposed model, and then re-estimated the nested models by incrementally increasing and decreasing the number of segments.

Table 2 Likelihood Function of Proposed and Nested Models

Model	Description	Likelihood Function For a Household	Number of Parameters
	Proposed model[1]	$\sum_{l=1}^{L} \pi_l \delta_{hj} \left(\frac{\lambda_{h\mid l}^{n_h}}{(e^{\lambda_{h\mid l}} - 1) n_h!} \right) + \sum_{s=1}^{S} \pi_s \left(\frac{n_h!}{\prod_j n_{hj}!} \prod_j p_{hj\mid s}^{n_{hj}} \right) \left(\frac{\lambda_{h\mid s}^{n_h}}{(e^{\lambda_{h\mid s}} - 1) n_h!} \right)$	$L(R + 1) + S(R + V + 1) - 1$
Nested Models			
N_1	No perceptual attributes[2]	Same as proposed model	$L(R + 1) + S(R + V^* + 1) - 1$
N_2	No loyal segments	$\sum_{s=1}^{S} \pi_s \left(\frac{n_h!}{\prod_j n_{hj}!} \prod_j p_{hj\mid s}^{n_{hj}} \right) \left(\frac{\lambda_{h\mid s}^{n_h}}{(e^{\lambda_{h\mid s}} - 1) n_h!} \right)$	$S(R + V + 1) - 1$
N_3	No loyal segments No choice segments	$\left(\frac{n_h!}{\prod_j n_{hj}!} \prod_j p_{hj}^{n_{hj}} \right) \sum_{s=1}^{S} \pi_s \left(\frac{\lambda_{h\mid s}^{n_h}}{(e^{\lambda_{h\mid s}} - 1) n_h!} \right)$	$S(R + 1) + V - 1$
N_4	No volume segments	$\frac{\lambda_h^{n_h}}{(e^{\lambda_h} - 1) n_h!} \left[\sum_{l=1}^{L} \pi_l \delta_{hj} + \sum_{s=1}^{S} \pi_s \left(\frac{n_h!}{\prod_j n_{hj}!} \prod_j p_{hj\mid s}^{n_{hj}} \right) \right]$	$R + S(V + 1) + J - 1$
N_5	No loyal segments No volume segments	$\frac{\lambda_h^{n_h}}{(e^{\lambda_h} - 1) n_h!} \left[\sum_{s=1}^{S} \pi_s \left(\frac{n_h!}{\prod_j n_{hj}!} \prod_j p_{hj\mid s}^{n_{hj}} \right) \right]$	$R + S(V + 1) - 1$
N_6	No segmentation	$\left(\frac{n_h!}{\prod_j n_{hj}!} \prod_j p_{hj}^{n_{hj}} \right) \left(\frac{\lambda_h^{n_h}}{(e^{\lambda_h} - 1) n_h!} \right)$	$R + V$

[1] L is the number of loyal segments, S is the number of switching segments, J is the number of brands, R is the number of volume-related covariates (including an intercept), and V is the number of choice-related covariates including $(J - 1)$ brand-specific constants.
[2] V^* denotes the reduced set of choice-related covariates.

the proposed model. Overall, TAE for N_1 is 10–30% higher than the TAE for the full model.[4] Further, under N_1 it is harder to interpret the switching segments. As we will discuss later, under the proposed model these segments have an intuitively appealing and managerially useful interpretation. In section 5 we elaborate on this model diagnostic by indicating how information about the perceptual attributes and consumer segments can help a company launch a new brand with minimal cannibalizing effects on its existing brand.

Nested Model 2 (N_2) estimates S segments, but does not distinguish between loyal and switching segments. Two versions of N_2 have been proposed elsewhere. Poulson (1983) proposed the basic multinomial-Poisson model which did not include covariates. More recently, Böckenholt (1993b) presents a multinomial-Poisson regression model that includes covariates in the context of analyzing recurrent choice data from scanner panels. Without loyalty segments, this reduces the number of parameters to almost half of the full model. However, both the CAIC (which adjusts for the number of parameters) and TAE suggest that the fit of N_2 is worse than the full model (TAE is 4–14% worse).[5] Two issues

[4] As expected, a more constrained nested model which includes no covariates (except intercepts), performs even worse with 15–46% worse TAE.

[5] The likelihood ratio test is not valid for models N_2–N_6 because when the structure of the mixture model is changed, the likelihood ratio is no longer asymptotically χ^2-distributed (McLachlan and Basford 1988, p. 21–31). In such cases, CAIC, which penalizes a model for estimating additional

Table 3 Comparison of Proposed, Nested, and Competing Models

		No. of Segments					Total Absolute Error			
							Total Sample		Holdout Sample	
Model	Description	Loyal	Switch	Number of Parameters	$\mathcal{LL}$	CAIC	Agg.	Disagg.	Agg.	Disagg.
	Proposed model	18	5	151	−4009	9329	135.2	446.3	78.7	194.6
Nested Models										
N_1	No perceptual attributes	18	5	126	−4753	10600	162.9	580.2	89.4	213.7
N_2	No loyal segments	0	5	79	−4482	9650	148.9	507.0	85.2	201.3
N_3	No loyal segments No choice segments	0	5	31	−5789	11969	174.5	763.2	150.1	300.7
N_4	No volume segments	7	5	74	−4398	9438	147.1	496.3	86.1	203.3
N_5	No loyal segments No volume segments	0	5	67	−4601	9784	154.1	517.1	88.2	203.7
N_6	No segmentation	0	0	15	−6384	12898	218.3	874.8	198.1	425.4
Competing Models										
C_1	Modified Grover & Srinivasan (1992)	7	4	115	−11519	24037	205.3	830.8	168.1	335.7
C_2	Modified Krishnamurthi & Raj (1991)	–	–	169	−10190	21848	214.1	859.9	208.6	469.7
C_3	Modified Colombo & Morrison (1989)	7	1	22	−5399	10989	168.8	551.9	140.3	289.7
C_4	NBD-Dirichlet	–	–	9	−6193	12464	210.9	807.2	172.1	356.7
C_5	NBD-Dirichlet with Hard-Core Loyals	7	1	16	−5978	12095	198.3	782.7	157.4	317.8

warrant notice. First, many previous studies have suggested separating the loyal and switchers (e.g., Colombo and Morrison 1989, Grover and Srinivasan 1992). Second, separating loyal and switchers may also be useful from an estimation standpoint. If the data include hard-core loyal consumers, but the model structure does not specify them as a separate group, an attempt to empirically separate out the loyal and switchers may lead to some parameter estimates of $\pm\infty$.

Nested Model 3 (N_3) recognizes category volume segments but assumes choice probabilities are invariant across segments. This reduces the number of parameters substantially. However, this also makes the model significantly worse than the full model—CAIC of N_3 is

parameters, is a more appropriate index. Lower the CAIC, better the model. Based on this criterion, the proposed model is better than N_2.

Table 4 · Observed versus Estimated Loyalty Share

	% Loyal Households	
Brands	Observed	Estimated
Private Label	9.19	4.93
Bounty	11.90	10.14
Brawny	5.79	4.73
Scott	5.51	3.63
Viva	3.17	1.89
Hi-Dri	3.49	2.61
Mardi Gras	1.93	1.14
Total	40.97	29.09

much larger, and TAE is 30–90% greater than that for the proposed model.

Nested Model 4 (N_4) only recognizes loyal and switching segments with respect to brand choice. Volume-related parameters are assumed invariant.

Nested Model 5 (N_5) further constrains N_4 by assuming no loyal segments. CAIC for these nested models suggest that the full model is superior even after it is penalized for estimating additional parameters. TAE for N_4 and N_5 are respectively 5–11% and 5–16% worse than that for the full model. This suggest that volume segmentation is useful. Comparing N_3 and N_4 we find that at least in this application, choice segmentation is more important than volume segmentation.

Nested Model 6 (N_6) represents the most parsimonious model where category volume and brand choice probability parameters are assumed to be invariant across segments. In other words, this model is consistent with a single-segment (or no segmentation) model. This model estimates only 15 parameters. However, this parsimony comes at a heavy cost as shown by a dramatic drop in model fit and predictive ability. For example, TAE for N_6 is 61–152% worse than the TAE for the proposed model. Interestingly, the model with segments but no covariates performs much better than the model with covariates but no segments. In other words, in our empirical application, market segmentation is more important than the inclusion of covariates.

4.4.2. Competing models. Five competing models are estimated to evaluate the performance of the proposed model.

Competing Model 1 (C_1) is a modified version of Grover and Srinivasan ((G&S) 1992) model. This model consists of three steps. First, we use the *observed brand share* for each household, to obtain loyal and switching groups, based on cluster analysis (*K*-Means) and iterative Bayesian procedure. In our application, we obtain 7 loyal and 4 switching segments. Second, for each switching segment, we fit a separate logit choice model. Although G&S fit an *aggregate* logit model, we chose to fit a disaggregate (or household) level logit model in order for this model to be more comparable to our model.[6] Third, for each loyal and switching segment obtained from the first step, we estimate a category volume model using a simple regression. Once again, we use a disaggregate or household-level volume model instead of the aggregate volume model used by G&S. Across all three steps, the total $\mathcal{LL}$ is -11519 and the number of estimated parameters is 115 (see Table 3). Notice that clustering in step one is not a parametric method. Therefore, it is unclear whether the $\mathcal{LL}$ and "parameters" from this step should be included in model fit statistics. Perhaps, TAE is a better comparative measure. As indicated in Table 3, TAE for C_1 is 52–114% worse than the proposed model. Notice that C_1 does not segment consumers on the basis of their category volume. Comparing the TAE of C_1 with that for nested models N_4 and N_5 (which also estimate only choice segments), we find that the performance of models N_4 and N_5 is superior to that of C_1. This suggests that segmenting consumers on the basis of their response to covariates (as done in N_4 and N_5) is better than segmenting consumers on the basis of their observed brand share (as done in C_1).

Competing Model 2 (C_2) is a modified version of Krishnamurthi and Raj ((K&R) 1991) model. This model also consists of three steps. First, household h is defined as loyal to brand j ($LOY_{hj} = 1$), if the observed share of brand j for that household is greater than 0.5, otherwise $LOY_{hj} = 0$. In our application, 62.9% of the households were classified as loyal to one of the 7 brands. Second, we estimate a single disaggregate logit model for all households where the choice utility is given by

[6] Disaggregate models are generally considered better than aggregate models because they utilize the rich household level information.

$$U_{hj} = a_{j0} + \sum_{v=1}^{V} a_{jv} x_{jv} + b_j \text{ LOY}_{hj} + \sum_{v=1}^{V} c_{jv}(\text{LOY}_{hj} \times x_{jv}). \quad (23)$$

Effectively, this formulation specifies a main-effect of brand choice covariates (x_{jv}) and their interaction with the loyalty variable. This allows loyal households to have a different response to marketing variables than the non-loyal households. Finally, for each brand j, we run a separate disaggregate regression on brand quantity as:

$$\ln Q_{hj} = \beta_{j0} + \sum_{r=1}^{J} \beta_{jr} \text{ Price}_{hr} + \gamma_{j1} \text{ LOY}_{hj} + \gamma_{j2}(\text{LOY}_{hj} \times \text{Price}_{hj}) + \delta_1 \text{ DUM}_h + \delta_2 FS_h \quad (24)$$

where $\text{DUM}_h = 1$ is household h is a heavy buyer of the category (above median), 0 otherwise; and FS_h is the family size of household h (defined as the number of children in our application). C_2 estimates a total of 169 parameters with a $\mathcal{LL}$ of -10190, and CAIC of 21848. Further, the TAE for C_2 are 58–141% worse than those for the proposed model. These statistics suggest that the proposed model performs significantly better than C_2. Comparing the TAE for C_1 and C_2, we find that C_1 performs better than C_2.

Competing Model 3 (C_3) is a modified version of Colombo & Morrison ((C&M) 1989). This model assumes one hard-core loyal segment for each brand and one switching segment with a zero-order choice process. Although the C&M model did not include either the category volume or the covariates, we include them in this modified version to make it more comparable to our model. Further, C&M use switching data to estimate their model. However, since their model assumes a zero-order process, it is possible to estimate this model using frequency data. The likelihood for this model is:

$$\mathcal{L}(\mathbf{y}_h) = \frac{\lambda_h^{n_h}}{(e^{\lambda_h} - 1)n_h!} \left[\sum_{j=1}^{J} \pi_j \delta_{hj} + \pi_s \left(\frac{n_h!}{\prod_j n_{hj}!} \prod_j p_{hj|s}^{n_{hj}} \right) \right] \quad (25)$$

where $p_{hj|s}$ is defined in (6), π_j is the size of the segment loyal to brand j, π_s is the size of the switching segment, and $\Sigma_j \pi_j + \pi_s = 1$. Results show that although C_3 performs significantly better than C_1 and C_2, it is substantially worse than the proposed model (TAE is 25–78% worse).

Competing Model 4 (C_4) is the NBD-Dirichlet model. This model assumes that category purchases of a household are distributed Poisson, and heterogeneity in household purchase rates can be captured by Gamma distribution. The brand choice model is a mixture of multinomial model at the household level with Dirichlet heterogeneity. The choice and volume models are assumed to be independent of each other (Goodhardt et al. 1984). Appendix 2 gives the likelihood function for this model. This model includes no covariates and estimates only 9 parameters. However, all the statistics of Table 3 suggest that this model performs significantly worse than the proposed model. Interestingly, the model fits and predicts almost as well as C_1 and C_2 with less than 1/10th the number of parameters.

Competing Model 5 (C_5) is the NBD-Dirichlet model with one hard-core loyal segment for each brand. The likelihood function of this model is also given in Appendix 2. Although this model estimates more parameters than NBD-Dirichlet model, the CAIC and TAE suggest that it is better to include loyal segments in the model. These statistics also indicate that the proposed model is significantly better than C_5.

In summary, this analysis suggests that the proposed model is significantly better than all the nested and competing models. Table 3 provides a good indication of incremental contribution of each model component.[7]

4.5. Results for the Proposed Model

4.5.1. Loyal segments. Table 4 gives the percentage of households who are *observed* and *estimated* to be loyal to each brand. If a household reports buying the same brand in the last 4 weeks, then that household is *observed loyal* to that brand. However, as discussed earlier, this does not necessarily mean that this household is truly loyal to that brand since it may have bought the same brand several times in a row just by chance. The

[7] We thank an anonymous reviewer and the Area Editor for suggesting these comparisons.

estimated loyalty share corrects for this chance occurrence and will therefore be always less than observed loyalty share. Table 4 results confirm this.

Each estimated loyal segment is further divided into sub-segments based on the category purchase volume of its consumers. Table 5 provides the parameter estimates, segment shares, and segment profile of these category volume segments. Five major conclusions can be drawn from these results. First, all the significant coefficients in Table 5 have the expected signs.

Second, for all the brands, category volume segments emerge on the basis of consumers' price sensitivity. For example, Bounty has one segment of loyal consumers who are not price sensitive in their category volume behavior (β not significant), another segment where consumers are somewhat price sensitive ($\beta = -0.57$), and a third segment where consumers are highly price sensitive ($\beta = -1.21$). This pattern is consistent across all brands.

Third, in general the price sensitive segments (especially the highly price sensitive segments) also have a significant positive coefficient for the number of children. Combining this result with the average number of children in the household and average purchase rate of each segment (as given under segment profile), we see that households with more children are heavier users of paper towels and that their demand for paper towels category is more price sensitive. This is consistent with our expectations and industry speculations.

Fourth, in almost all the high price sensitive segments of the 7 brands (e.g., segment 3 for private label, Bounty, Brawny etc.) the constant term for the category volume model is not significant. This suggests that almost all the variance in these segments is captured by the two variables, price and number of children.

Fifth, all brands, except Hi-Dri and Mardi Gras, have at least one price insensitive segment. However, across the 7 brands, only 9.72% (1.36% of Private Label + 3.21% of Bounty + $\cdots$) of the loyal households are price insensitive in their volume decision.

These results suggest that the loyal households show increasing price sensitivity in their category volume decision if they have more children and are heavier users of paper towels.

4.5.2. Switching segments. Our model estimates that about 71% or 1720 households are switchers. These

Table 5 Loyal Segments for Paper Towel

Brands	Private Label			Bounty			Brawny			Scott		Viva			Hi-Dri		Mardi Gras	
Category Segments	1	2	3	1	2	3	1	2	3	1	2	1	2	3	1	2	1	2
Category Volume Model:																		
Parameters & (*t*-values)[1]																		
Constant	0.79	1.86	−0.10	0.48	0.36	0.09	0.45	0.74	−0.50	−0.12	0.77	0.35	0.85	−1.01	0.58	0.66	0.84	−0.45
	(4.11)	(4.38)	(−0.98)	(2.22)	(2.66)	(0.71)	(6.92)	(6.27)	(−1.11)	(−1.69)	(1.99)	(3.15)	(2.54)	(−2.45)	(3.07)	(1.66)	(2.22)	(−1.97)
Price	−0.84	−1.90	−2.71	0.11	−0.57	−1.21	0.17	−0.09	−0.95	−0.13	−0.71	−0.11	−0.99	−1.69	−0.52	−1.01	−0.88	−1.47
	(−1.70)	(−3.58)	(−8.81)	(0.57)	(−3.63)	(−8.18)	(0.34)	(−0.20)	(−4.48)	(−1.13)	(−3.27)	(−1.06)	(−3.06)	(−12.16)	(−2.02)	(−4.37)	(−2.94)	(−7.00)
# Children	0.54	0.68	1.41	0.12	0.88	1.13	0.22	0.41	1.32	0.66	0.60	0.31	0.74	1.51	0.32	0.68	0.61	1.31
	(1.19)	(1.55)	(9.16)	(0.48)	(2.56)	(5.38)	(1.11)	(1.15)	(5.71)	(2.53)	(3.29)	(0.94)	(3.30)	(7.33)	(0.68)	(2.31)	(1.16)	(10.40)
Segment Share (%)	1.36	2.56	1.02	3.21	3.65	3.28	1.13	1.70	1.89	1.85	1.80	0.47	0.59	0.83	0.83	1.78	0.48	0.66
Segment Profile:																		
Price ($)	0.68	0.60	0.52	0.96	0.91	0.84	0.83	0.82	0.74	0.86	0.80	0.90	0.87	0.52	0.68	0.65	0.80	0.74
# Children	1.3	0.9	2.2	1.3	1.5	2.0	1.0	1.0	1.9	1.4	1.6	1.1	1.5	2.2	1.3	1.6	1.3	2.1
Purchase Rate (# rolls/month)	2.5	3.8	4.9	2.1	3.2	3.8	2.2	3.3	3.7	2.0	3.0	1.8	3.0	4.2	1.9	3.0	2.5	3.4

[1] *t*-values in parentheses.

consumers can be grouped into 5 segments based on their choice and volume decisions (Table 6).

Segment 1 (24% share) is extremely price sensitive in both choice and volume. In the brand choice model, price is highly significant. Further, of the 5 perceptual attributes only all-purpose and economy are significant, with economy being the dominant factor. Therefore we refer to this segment as the *Price Segment*. In the category volume model, both price and number of children are significant and are in the expected direction. The profile of this segment shows that on average households in this segment have more children and are heavier users of paper towels. This explains their high price sensitivity in both choice and volume decisions. Not surprisingly, private labels have a dominant share (30%) in this segment (see Table 7).

Figure 1 gives good insight about the relative position of brands in segment 1. Here we use the two significant perceptual attributes for this segment: all-purpose and economy. Notice two things. First, brand positions on the perceptual attributes and brand shares show a remarkably good relationship. Second, Bounty and Viva are positioned next to each other while private label is positioned far away from these brands. This suggests that if Bounty offers a price discount, it will probably hurt Viva's sales more than that of private label brands. We will show in Section 5, that this observation is consistent with our results of the price elasticity analysis.

Table 6 Switching Segments for Paper Towel

Segments	1	2	3	4	5
Brand Choice Model (Parameters & *t*-values):[1]					
Strength	0.85	5.12	0.66	0.87	2.88
	(0.95)	(17.01)	(2.21)	(1.40)	(3.29)
Absorbing	0.78	1.87	6.33	1.11	3.28
	(1.13)	(4.21)	(7.90)	(1.23)	(5.48)
Tearing	−1.66	−2.78	−1.89	5.32	0.30
	(−1.18)	(−3.13)	(−1.06)	(15.42)	(0.92)
All Purpose	1.88	1.11	3.15	0.55	6.06
	(2.52)	(1.36)	(3.46)	(1.27)	(12.17)
Economical	7.12	0.99	−0.32	0.56	4.12
	(17.76)	(0.99)	(−0.69)	(1.30)	(4.12)
Price	−5.39	−1.42	−0.88	0.17	−3.90
	(−7.25)	(−3.54)	(−1.69)	(0.52)	(−4.46)
Category Volume Model (Parameters & *t*-values):					
Constant	−1.26	1.42	0.58	−0.44	0.56
	(−3.78)	(3.04)	(2.90)	(−1.30)	(5.00)
Price	−2.19	−0.81	−1.01	−0.15	−2.05
	(−16.34)	(−3.93)	(−5.58)	(−0.94)	(−9.72)
No. of Children	1.91	0.51	0.98	0.76	1.55
	(13.45)	(3.49)	(3.28)	(2.41)	(9.06)
Segment Share (%)	23.79	15.47	8.11	3.11	20.42
Segment Profile:					
Price ($)	0.79	0.90	0.86	0.91	0.80
No. of Children	2.4	1.1	1.5	1.2	1.8
Purchase rate (rolls/month)	4.9	3.5	3.3	1.4	5.5

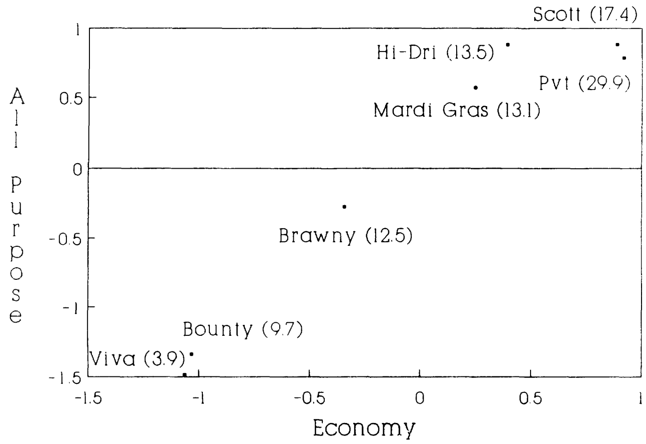
[1] *t*-values in parentheses.

Table 7 Within-Segment Brand Share[1]

Brand	Price Segment	Strength Segment	Absorbency Segment	Tearing Ease Segment	Value Segment
Private Label	29.93	5.91	2.72	7.23	28.11
Bounty	9.65	40.08	33.99	7.37	17.14
Brawny	12.53	8.34	28.39	29.50	11.24
Scott	17.38	26.78	7.25	7.37	15.02
Viva	3.86	15.32	9.86	24.78	18.89
Hi-Dri	13.52	2.97	15.52	5.90	7.35
Mardi Gras	13.13	0.60	2.27	17.85	2.25

[1] Three brands with the highest share within each segment are underlined.

Figure 1 Perceptual Map for Segment 1



Segment 2 (15.5%) shows moderate price sensitivity in both choice and volume. This segment values strength and absorbency, with strength being the most dominant perceptual attribute. Therefore we refer to this segment as the *Strength Segment*. Table 7 shows that Bounty is the dominant player (40% share) in this segment. This is consistent with the strength and absorbency positioning of Bounty. Although Brawny is also positioned as a strength brand (recall the tough guy on Brawny pack), it is not doing very well in this segment. In fact Scott is doing much better than Brawny. It is interesting to compare these results with a recent *Consumer Reports* (January 1992) evaluation of several brands of paper towels on certain objective criteria. For example, *Consumer Reports* evaluated strength of a paper towel by determining how much lead shot a wet towel could support and how well it held up during scrubbing. Based on this criterion, on a 1–5 scale, Bounty and Scott were rated 4, Brawny and Viva were rated 3, Hi-Dri, and private labels were rated 2, and Mardi Gras obtained a rating of 1. This ordering of ratings is quite consistent with the brand shares in the strength segment (Spearman rank correlation = 0.97).

Segment 3 (8%) is price insensitive in choice and moderately price sensitive in volume. This segment looks for paper towels with strength, absorbency, and all-purposeness. Absorbency is the dominant attribute in this segment. Therefore we refer to this segment as the *Absorbency Segment*. The volume model and the profile of this segment are not very different from that of segment 2, i.e. moderate purchase rate and small family size. Bounty, Brawny, and Hi-Dri are doing well in this segment. Once again the brands with high share in this segment also get high ratings on absorbency in *Consumer Reports* (Spearman rank correlation = 0.66). Interestingly, James River Corp. recently decided to reposition Brawny by replacing its 10-year-old tag line "The Big, Tough Towel" with "Thirst Pockets for Spill Relief" (*Marketing News*, April 11, 1994). This shift in

Table 8 Sales Increase for Bounty for a 5% Drop in Its Price

Segments	Bounty Sales Before Price Cut	Bounty Sales After Price Cut	% Increase in Sales Due to Category Volume	Brand Switching	Total
Loyal Segments for Bounty					
1	147	154	4.7	0.0	4.7
2	254	270	6.3	0.0	6.3
3	272	300	10.3	0.0	10.3
All Loyals	673	724	7.6	0.0	7.6
Switching Segments					
1	245	365	20.4	28.6	49.0
2	473	489	0.6	2.8	3.4
3	198	212	4.5	2.6	7.1
4	7	7	0.0	0.0	0.0
5	419	494	6.4	11.5	17.9
All Switchers	1342	1566	6.6	10.1	16.7
Total	2015	2290	6.9	6.7	13.6

Brawny's position is consistent with our results which show that Brawny has a substantially higher share in the Absorbency segment (28%) than in the Strength segment (8%).

Segment 4 (3%) is the smallest switching segment. Consumers in this segment are price insensitive in both their choice and volume decisions. The brand choice decision of these consumers is affected by the tearing ease of paper towels. Therefore we refer to this segment as the *Tearing Ease Segment*. These households have a small family size and are light users of paper towels, which explains their price insensitivity. Brawny, Viva and Mardi Gras are the major brands in this segment. This is consistent with the high *Consumer Reports* ratings of these brands on tearing ease (Spearman rank correlation = 0.72).

Segment 5 (20.5%) is the second largest switching segment after segment 1. This segment is also very price sensitive in both choice and volume, though its price sensitivity is somewhat less than that of segment 1. Four of the five perceptual attributes are significant in this segment. Although all-purpose is the most dominant factor, all other attributes also have a reasonable impact on brand choice. In other words, consumers in this segment seem to be looking for a little bit of everything at a reasonable price. Therefore we refer to this segment as the *Value Segment*. The volume model for this segment is similar to that of segment 1, with high price sensitivity and significantly positive impact of number of children. This segment consists of moderate size families who are the heaviest users of paper towels. Once again private label emerges as a dominant player in this segment.

In summary, five switching segments emerge with distinct characteristics (price, strength, etc.). Brands with dominant share in a segment also have high objective ratings on that segment characteristic, as reported by *Consumer Reports*.

5. Price Elasticity Analysis

A brand manager needs to understand the impact of marketing mix variables on the sales of his own and competing brands. For example, the brand manager of Bounty may wish to know the increase in Bounty's sales if its price is reduced by 5%. Further, to anticipate competitive reaction, s/he may also want to know where this sales increase comes from. In this section we address these issues by using our segment level results. Specifically, we reduce Bounty's price by 5%, simulate category sales and brand shares in all segments, estimate brand sales before and after price cut, and compute gain or loss in brand sales.

Our results show that when Bounty reduces its price by 5%, its sales increase by 13.6%. A key advantage of our modeling approach is that we also get an estimate of the sources of such sales increases. Table 8 decomposes the 13.6% sales increase into various components. Of the total sales increase for Bounty (13.6%), almost half comes from brand switching (6.7%) and the other half from increase in volume (6.9%). We can not say with certainty if this volume increase is due to increase in consumption or if it is due to stockpiling by consumers. Given the nature of the product category the latter seems more likely. The total volume increase of 6.9% can be further decomposed into 2.5% increase from consumers loyal to Bounty, and 4.4% increase from switchers. Among the three loyal segments of Bounty, the largest percentage increase in sales comes from segment 3 which is also the most price sensitive segment. Among the switching segments the largest percentage increase in sales comes from segment 1 which is the most price sensitive switching segment. Switching segment 4 is price insensitive and hence it does not contribute to increase in Bounty's sales.

Table 8 indicates that 6.7% of Bounty's sales increase comes from competing brands. To gain a better understanding of this competitive structure, in Table 9 we

Table 9 Impact of Bounty's 5% Price Cut On Competitive Brands

		% Decrease in Sales of					
Switching Segment	Segment Share	Private Label	Brawny	Scott	Viva	Hi-Dri	Mardi Gras
1	23.8	0.9	4.3	3.6	7.5	1.5	6.1
2	15.5	2.4	1.5	1.7	1.7	2.1	2.9
3	8.1	6.7	0.3	0.9	2.2	0.9	5.6
4	3.1	0.0	0.0	0.0	0.0	0.0	0.0
5	20.4	0.8	2.3	5.1	2.2	1.0	9.8
Total	74.0	1.0	2.5	3.4	2.7	1.3	6.3

present the percentage loss in sales of competing brands due to a 5% drop in Bounty's price.[8] Some interesting observations emerge from this table. Bounty's price cut has the smallest impact on private labels and the largest effect on Mardi Gras. In the Price Segment (switching segment 1), Viva is hurt the most. This is not surprising given Viva's proximity to Bounty in the perceptual map (see Figure 1).

Notice from Tables 8 and 9 that Bounty gets its greatest sales increase from the two largest switching segments: segment 1 or the Price segment (24% share) and segment 5 or the Value segment (20% share). Even though private labels are the dominant brands in these two segments (see Table 7), they lose the least amount of sales to Bounty (see Table 9). This suggests that Procter and Gamble (manufacturer of Bounty) may wish to introduce a new brand of paper towel that priced and positioned directly against the private labels. Not only will this new brand have the possibility of getting share in two of the largest switching segments, but it will also have the least cannibalizing effect on Bounty.

Overall these results provide a good understanding of the effect of price cut on own brand sales, on category volume and switching, and on the sales of competing brands. Such an analysis offers very useful diagnostics to a brand manager.

6. Conclusion

The primary purpose of this research was to develop a model that can help brand managers better understand the sources of gain or loss for their brand sales. We focus on two distinct components which may affect the sales of a brand: category volume and brand share. By explicitly modeling these two components, our model can provide useful diagnostics. Sales loss due to switching can also be traced to relative gains by competing brands thereby providing a better understanding of the competitive structure in the market place. We also recognize consumer heterogeneity in response to marketing activity. Therefore all the analysis is done at the segment level. Segments are inferred based on consumers' response to marketing activities for *both* their brand choice *and* category volume decisions.

We provided a real-life application of our model for the paper towels category. We found that in general households showed increasing price sensitivity in their category volume decisions if they had more children and were heavier users of paper towels. Two of the largest switching segments (labeled as Price and Value Segments) were found to be very price sensitive in both their brand choice and category volume components. In both these segments, private label brands were dominant with about 30% share. Segments also emerged on the basis of Strength, Absorbency, and Tearing Ease of paper towels. Interestingly, we found that brand shares within these segments were quite consistent with the objective quality ratings of brands as given by *Consumer Reports*. Finally, price elasticity analysis for Bounty revealed that a 5% drop in Bounty's price increased its sales by 13.6%. Almost half of this increase came from brand switching, and the other half came from increase in category volume (e.g., stockpiling). Bounty gained the most from the price sensitive segments (Price and Value Segments). Private labels were hurt the least by Bounty's price cut. These results also provide useful guidelines for new product introductions.

We hope this research inspires marketing managers to go beyond the aggregate measures of brand sales, and look for the sources of sales gain or loss. Our model provides one approach to achieve this objective.[9]

[8] In the spirit of cross price elasticities, the decrease in sales of a competing brand is represented as a percentage of the sales of that brand before Bounty's price cut.

[9] The authors thank the Editor, the Area Editor, two anonymous reviewers, and the participants of Marketing Modelers Group, Columbia-Wharton-Yale colloquium, Carnegie-Mellon University and University of Washington, St. Louis seminar for their comments. Thanks are also due to an anonymous company for providing the data, and Narendra Mulani of IRI for providing supporting data.

Appendix 1: First and Second Derivatives

The Category Volume Model Component

The first and second derivatives of the "complete" data log-likelihood associated with the category volume (Poisson) component for segment k is given by

$$\frac{\partial \mathcal{L}\mathcal{L}_p}{\partial \gamma_{r|k}} = \sum_{h=1}^{H} \alpha_{k|\mathbf{y}_h} z_{hr} \left(n_h - \frac{\lambda_{h|k}}{1 - \exp(-\lambda_{h|k})} \right), \quad (r = 1, 2, \ldots, R)$$

$$\frac{\partial^2 \mathcal{L}\mathcal{L}_p}{\partial \gamma_{r|k} \partial \gamma_{i|k}} = -\sum_{h=1}^{H} \alpha_{k|\mathbf{y}_h} \lambda_{h|k} z_{hr} z_{hi} \left(\frac{1 - (1 + \lambda_{h|k}) \exp(-\lambda_{h|k})}{[1 - \exp(-\lambda_{h|k})]^2} \right),$$
$$(r, i = 1, 2, \ldots, R).$$

The Brand Choice Component

The first and second derivatives of the "complete" data log-likelihood associated with the brand choice (multinomial) component for segment k is given by

$$\frac{\partial \mathcal{LL}_{\mathcal{M}}}{\partial \beta_{v|k}} = \sum_{h=1}^{H} \alpha_{k|\mathbf{y}_h} \sum_{j'=1}^{J} y_{hj'} \left(x_{hj'v} - \sum_{j=1}^{J} x_{hjv} p_{hj|k} \right), \quad (v = 1, 2, \ldots, V)$$

$$\frac{\partial^2 \mathcal{LL}_{\mathcal{M}}}{\partial \beta_{v|k} \partial \beta_{m|k}} = -\sum_{h=1}^{H} \alpha_{k|\mathbf{y}_h} \sum_{j'=1}^{J} y_{hj'} \times \left\{ \left(x_{hj'v} - \sum_{j=1}^{J} x_{hjv} p_{hj|k} \right) \times \left(x_{hj'm} - \sum_{j=1}^{J} x_{hjm} p_{hj|k} \right) - \left(\sum_{j=1}^{J} p_{hj|k} (x_{hj'v} - x_{hjv})(x_{hj'm} - x_{hjm}) \right) \right\},$$

$$(v, m = 1, 2, \ldots, V).$$

Appendix 2: NBD-Dirichlet Model

1. NBD Model

The category volume can be captured by the NBD model. This model assumes that individual purchases are distributed Poisson, and the heterogeneity in consumers' purchase rates can be represented by a Gamma distribution (Morrison and Schmittlein 1988):

Poisson individual purchases: $\Pr_{\mathcal{P}}(y_h = n_h; \lambda_h) = \dfrac{\lambda_h^{n_h} e^{-\lambda_h}}{n_h!}$

Gamma heterogeneity: $g(\lambda_h; r, \alpha) = \dfrac{\alpha^r \lambda_h^{r-1} e^{-\alpha\lambda_h}}{\Gamma(r)}$

Unconditional probability:

$$\Pr_{\mathcal{N}}(\mathbf{y}_h; n_h; r, \alpha) = \binom{n_h + r - 1}{n_h} \left(\frac{\alpha}{\alpha + 1}\right)^r \left(\frac{1}{\alpha + 1}\right)^{n_h}$$

where λ_h is the purchase rate of household h, n_h are the total category purchases for this household, and r and α are parameters of the Gamma distribution, such that $r, \alpha > 0$.

2. Dirichlet Model

The brand choice can be captured by the Dirichlet model which assumes consumer purchases across brands are distributed multinomial and heterogeneity in consumers' purchase probability is accounted for by a Dirichlet distribution.

Multinomial individual purchases: $\Pr_{\mathcal{M}}(\mathbf{y}_h; n_h, \mathbf{p}_h) = \dfrac{n_h!}{\prod_j n_{hj}!} \prod_j p_{hj}^{n_{hj}}$

Dirichlet heterogeneity: $f(\mathbf{p}; \mathbf{a}_j) = \dfrac{\Gamma(S)}{\prod_{j=1}^{J} \Gamma(a_j)} \prod_{j=1}^{J} p_j^{a_j - 1}$

Unconditional choice probability:

$$\Pr_{\mathcal{D}}(\mathbf{y}_h; n_h, \mathbf{a}_j) = n_h! \frac{\Gamma(S)}{\Gamma(S + n_h)} \prod_j \frac{\Gamma(a_j + n_{hj})}{n_{hj}! \Gamma(a_j)}$$

where n_{hj} are the number of brand j purchases for household h, and a_j are the parameters of the Dirichlet distribution, such that $a_j > 0$, and $\Sigma_j a_j = S$.

3. NBD-Dirichlet Model

Assuming independence between the volume and choice models, the NBD and the Dirichlet model can be combined to arrive at the following (Goodhardt et al. 1984):

$$\Pr_{N-D}(\mathbf{y}_h; n_h, r, \alpha, \mathbf{a}_j) = \binom{n_h + r - 1}{n_h} \left(\frac{\alpha}{\alpha + 1}\right)^r \left(\frac{1}{\alpha + 1}\right)^{n_h} \frac{n_h! \Gamma(S)}{\Gamma(S + n_h)} \prod_j \frac{\Gamma(a_j + n_{hj})}{n_{hj}! \Gamma(a_j)}.$$

Using the properties of the Gamma function, i.e. $\Gamma(x) = (x - 1)\Gamma(x - 1)$, it can be easily shown that the log-likelihood function of this model is:

$$\mathcal{LL} = \sum_h \left[\sum_{k=0}^{n_h - 1} \ln(r + k) - \sum_{j=1}^{J} \sum_{k=1}^{n_{hj}} \ln k + r\{\ln \alpha - \ln(\alpha + 1)\} - n_h \ln(\alpha + 1) - \sum_{k=0}^{n_h - 1} \ln(S + k) + \sum_{j=1}^{J} \sum_{k=1}^{n_{hj}} \ln(a_j + k) \right].$$

For J brands, this model estimates a total of $J + 2$ parameters (r, α, and J of the a_j's).

4. NBD-Dirichlet Model with Hard-Core Loyals

The NBD-Dirichlet model can be modified to accommodate hard-core loyal consumers for each brand. The category volume model remains as NBD, and the brand choice model is now modified to have one loyal segment per brand, and one switching segment as per the Dirichlet model. The likelihood for this model can therefore be written as:

$$\mathcal{L} = \prod_h \left[\binom{n_h + r - 1}{n_h} \left(\frac{\alpha}{\alpha + 1}\right)^r \left(\frac{1}{\alpha + 1}\right)^{n_h} \right] \times \left[\sum_{j=1}^{J} \pi_j \delta_{hj} + \pi_s \frac{n_h!}{\prod_j n_{hj}!} \cdot \frac{\Gamma(S)}{\prod_j \Gamma(a_j)} \frac{\prod_j \Gamma(a_j + n_{hj})}{\Gamma(S + n_h)} \right]$$

where π_j is the size of the segment that is loyal to brand j, and π_s is the size of the switching segment such that $\Sigma_j \pi_j + \pi_s = 1$. The log-likelihood can therefore be written as:

$$\mathcal{LL} = \sum_h \left[\sum_{k=0}^{n_h - 1} \ln(r + k) - \sum_{j=1}^{J} \sum_{k=1}^{n_{hj}} \ln k + r\{\ln \alpha - \ln(\alpha + 1)\} - n_h \ln(\alpha + 1) + \ln\left\{ \sum_{j=1}^{J} \pi_j \delta_{hj} + \pi_s \frac{n_h!}{\prod_j n_{hj}!} \cdot \frac{1}{\prod_{k=0}^{n_h - 1} (S + k)} \cdot \prod_j \prod_{k=0}^{n_{hj} - 1} (a_j + k) \right\} \right].$$

This model has $2(J + 1)$ parameters.

References

Ben-Akiva, Moshe and Steven R. Lerman (1985), *Discrete Choice Analysis*. Cambridge, MA: MIT Press.

Berndt, E. K., B. H. Hall, R. E. Hall, and J. A. Hausman (1974), "Estimation and Inference in Non-Linear Structural Models," *Annals of Economic and Social Measurement*, 3, 653–665.

Böckenholt, Ulf (1993a), "Estimating Latent Class Distributions in Recurrent Choice Data," *Psychometrika*, 58, no. 3, 489–509.

——, (1993b), "A Latent Class Regression Approach for the Analysis of Recurrent Choice Data," *British Journal of Mathematical and Statistical Psychology*, 46, 95–118.

Brännäs and G. Rosenqvist (1994), "Semiparametric Estimation of Heterogeneous Count Data Models," *European Journal of Operational Research: Special Issue on Stochastic Models of Consumer Behavior*, forthcoming.

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

Bozdogan, Ham (1987), "Model Selection and Akaike's Information Criterion (AIC): The General Theory and its Analytical Extensions," *Psychometrika*, 52, 345–370.

Card, D. and D. Sullivan (1988), "Measuring the Effect of Subsidized Training Programs on Movements in and out of Employment," *Econometrica*, 56, 497–530.

Colombo, Richard A. and Donald G. Morrison (1989), "A Brand Switching Model with Implications for Marketing Strategies," *Marketing Science*, 8 (Winter), 89–99.

Dempster, A. P., N. M. Laird, and D. B. Rubin (1977), "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society*, B39, 1–38.

Dzhafarov, Ehtibar N. and Ulf Böckenholt (1993), "Decomposition of Recurrent Choices into Stochastically Independent Counts," Working Paper, University of Illinois at Urbana-Champaign.

Goodhardt, Gerald J., Andrew S. C. Ehrenberg, and Christopher Chatfield (1984), "The Dirichlet: A Comprehensive Model of Buying Behavior," *Journal of the Royal Statistical Society*, Series A, 147, 621–655.

Goodstein, Ronald C. (1994), "UPC Scanner Pricing Systems: Are They Accurate?," *Journal of Marketing*, 58 (April), 20–30.

Grover, Rajiv and V. Srinivasan (1992), "Evaluating the Multiple Effects of Retail Promotions on Brand Loyal and Brand Switching Segments," *Journal of Marketing Research*, 29 (February), 76–89.

Gupta, Sunil (1988), "Impact of Sales Promotions on When, What, and How Much to Buy," *Journal of Marketing Research*, 25 (November), 342–355.

Heckman, James J. and Robert J. Willis (1977), "A Beta-Logistic Model for the Analysis of Sequential Labor Force Participation by Married Women," *Journal of Political Economy*, 85, 27–58.

Kamakura, Wagner A. and Gary J. Russell (1989), "A Probabilistic Choice Model for Market Segmentation and Elasticity Structure," *Journal of Marketing Research*, 26 (November), 379–390.

Krishnamurthi, Lakshman and S. P. Raj (1988), "A Model of Brand Choice and Purchase Quantity Price Sensitivities," *Marketing Science*, 7 (Winter), 1–20.

—— and —— (1991), "An Empirical Analysis of the Relationship Between Brand Loyalty and Consumer Price Elasticity," *Marketing Science*, 10 (Spring), 172–183.

Kruger, Michael E. (1990), "Scanner Data Quality Issues: An Update and Implications for Decision Making," Presentation at the ARF Research Quality Workshop.

Lazarsfeld, Paul F. and Neil W. Henry (1968), *Latent Structure Analysis*. NY: Houghton-Miffin Company.

Lindsay, B. G. (1983), "The Geometry of Mixture Likelihoods," *The Annals of Statistics*, 11, 86–94.

McLachlan, Geoffrey J. and Kay E. Basford (1988), *Mixture Models: Inference and Applications to Clustering*. NY: Marcel Dekker.

Morrison, Donald G. and David C. Schmittlein (1988), "Generalizing the NBD Model for Customer Purchases: What Are the Implications and Is It Worth the Effort?," *Journal of Business and Economic Statistics*, 6 (April), 145–160.

Poulson, Carsten S. (1983), *Latent Structure Analysis with Choice Modeling Applications*. Unpublished doctoral dissertation, University of Pennsylvania.

Simar, L. (1976), "Maximum Likelihood Estimation of a Compound Poisson Process," *The Annals of Statistics*, 4, 1200–1209.

Tellis, Gerald J. (1988), "Advertising Exposure, Loyalty, and Brand Purchase: A Two-Stage Model of Choice," *Journal of Marketing Research*, 25 (May), 134–144.

Titterington, D. M., A. F. Smith, and V. E. Makov (1985), *Statistical Analysis of Finite Mixture Distributions*. NY: Wiley.

Winer, Russell S. (1983), "Attrition Bias in Econometric Models Estimated With Panel Data," *Journal of Marketing Research*, 20 (May), 177–186.

This paper was received January 31, 1994, and has been with the authors 1 month for 1 revision; processed by Russell S. Winer, Area Editor.