

**A Product Choice Model  
with Marketing, Filtering and Purchase Feedback**

John D.C. Little  
Eric T. Anderson

M.I.T. Sloan School of Management  
Working Paper No. 3670-94  
10 April 1994

Revision of a paper presented at  
TIMS/ORSA Marketing Science Conference  
St. Louis, Missouri  
March 1993

Authors' address: Room E56-308  
M.I.T. Sloan School of Management  
Cambridge MA 02142

## Abstract

A household's purchase history provides evidence about three interrelated phenomena, *product loyalty*, *marketing response* and *purchase feedback*. First, most households display underlying product preferences or *loyalties*, in the sense of habitual purchase. However, *marketing* activities are likely to influence brand choice. Finally, through *feedback* of experience with purchased products, prior preferences may be reinforced and new ones introduced. The three phenomena are intertwined. For example, if a product is bought under promotional conditions, one can ask whether the household made the purchase because of an underlying preference or because of the promotion. The formulation presented here seeks to untangle these effects.

Our model assumes that, on a given purchase occasion, a household's probabilities for choosing products follow a Dirichlet distribution. Initially, the Dirichlet reflects the heterogeneity of preferences across the household population. The parameters of the Dirichlet are modeled as functions of marketing variables. After each purchase the standard update of parameters for a Dirichlet-multinomial process provides new information about underlying household preferences. The presence of the marketing model permits the filtering of marketing influences from the preferences. Next, purchase feedback increases the probability that the product just purchased will be purchased again. Finally, we consider that other, unknown disturbances may influence household choice and tend to bring it back toward market share norms.

A comparison of the model with those of Guadagni and Little (1983) and Fader and Lattin (1993) on data sets for juice drinks and detergents shows the new model to provide an improved fit in these cases. Purchase feedback turns out to be statistically and operationally significant in the calibrated models. As a result, current marketing activities create future value. Analytic formulas for special cases provide insight into the amount and duration of the future effects.

**Keywords:** choice models, brand choice, buyer behavior, estimation

## Introduction

For marketing models to be useful to a marketer, they should do more than just fit data closely or forecast well on hold out datasets. Although these criteria are obviously important, models can sometimes fit well without providing an understanding of the process being modeled. Ideally, model parameters should link rather directly to operational phenomena and have useful marketing interpretations.

With this in mind, we study a class of household-based choice models calibrated on scanner panel data. The arrival of single source data has made possible disaggregate product choice models that contain marketing and other environmental variables. Among the most popular has been the multinomial logit, first used on scanner data by Guadagni and Little (1983). These authors introduced variables, which they called *loyalties*, that turned out to be extremely influential for explaining purchase behavior. Guadagni and Little define loyalty as an exponentially smoothed function of past household purchases, treated as 0-1 variables. At each purchase occasion a household's loyalty is a vector of non-negative values summing to one. Loyalty to a brand may be thought of as that brand's share of the household's past purchases, with recent purchases receiving a greater weight than earlier ones.

As Guadagni and Little point out, the loyalty variables encompass two different phenomena. First, households have heterogeneous preferences. Loyalties reflect and reveal these underlying preferences in the population. Second, preferences may change over time. Loyalties, as defined, adapt to changes in each household's preferences in a purchase feedback process. That is, the loyalties shift in the direction of the most recent purchase. These two phenomena are mixed together by the loyalty variable since the calibration is both across households and over time and only one smoothing parameter is used.

Separation of the phenomena is critical for practice. If the smoothing constant primarily reflects purchase feedback, then an extra purchase due to marketing activity will increase loyalty and have valuable continuing effects into the future. Furthermore, the model can be used to calculate them. On the other hand, if the smoothing constant primarily represents the model's learning of the preferences of the household, then the primary value of the loyalty variable will lie in facilitating the estimation of response parameters rather than in forecasting.

Furthermore, as noted by Lattin (1991), another phenomenon is taking place. The presence of marketing activities, particularly promotions by retail stores through special displays, newspaper features and temporary price reductions, can cause a household to purchase products differently from its usual preferences. Therefore, measures of loyalty based on a simple smoothing of past purchases may not exactly represent a household's underlying preferences. Lattin suggests that we should adjust for such promotional effects, a process he calls *filtering*.

Thus we hypothesize three phenomena: (1) underlying household preferences that hold under standardized marketing conditions and are heterogeneous across the population, (2) purchases that differ from these prior preferences because of marketing and (3) purchase

feedback caused by experience with the products bought. The original Guadagni-Little loyalty variable does not discriminate among these phenomena.

Empirical research has taken several approaches to modeling heterogeneity and purchase feedback. Applications of the multinomial logit since Guadagni and Little (1983), have almost always used loyalty-like variables but strategies have differed widely. Samuels (1983) seeks to separate the cross-sectional heterogeneity of preference from learning through purchase feedback by using a static loyalty based on a preperiod and a dynamic deviation from it. The most common approach has been to try to capture the heterogeneity and ignore or downplay purchase feedback. For example, Krishnamurti and Raj (1988) use a "brand purchase share history." This is a brand's fraction of total purchases from the start of the data up to the current purchase occasion. Such an approach has the property that, if household purchase probabilities are stable, the variable will converge to those probabilities. However, since the measure becomes increasingly hard to move with new data as the time series becomes long, it stresses the learning of household preferences from history and downplays adaptive change. Pedrick and Zufryden (1991) have both a long and short term loyalty. Long term loyalty is represented by the number of purchases by the household in an initialization period. Short term loyalty is the number in the most recent time period. This approach contains both adaptation and underlying preference, although, as in Samuels' model, the preferences set by the initialization period are not allowed to change. Bucklin and Lattin (1992) use household share in an initialization period to capture heterogeneity in preferences and use last brand purchased to capture the time varying component of preference. None of these authors try to filter out marketing influences.

Other approaches to modeling heterogeneity include fixed and random effects models. Fully general fixed effects models would allow each household to have its own set of parameters. Few papers have explored this avenue because estimation of parameters is impractical without longer time series than are usually available. Rossi and Allenby (1993), however, have used Bayesian estimators to produce individual household estimates of multinomial logit parameters. Random effects models reduce the number of parameters by assuming that each household's parameters follow a distribution. The distribution may be specified by the researcher, in a parametric approach, or determined from the data, using semiparametric methods. Some recent applications of this approach are Chintagunta, Jain, and Vilcassim (1991) and Gönül and Srinivasan (1993) using logit models and Vilcassim and Jain (1991) with a semi-Markov formulation. While these models offer flexible specifications of heterogeneity, they contain no purchase feedback or filtering.

Recent progress in estimation has made procedures that require integration over multivariate normal distributions more computationally tractable than heretofore. Allenby and Lenk (1992) approach heterogeneity through random effects modeled in this manner. Their model introduces an autocorrelation of residuals that addresses phenomena related to our purchase feedback. Our model differs from theirs in many ways, as will be seen, and, in particular, we shall model the purchase feedback process with a specific, operationally motivated parameter.

The model we propose is logit based and has a similar flavor to the parametric random effects models with respect to heterogeneity in that we assume a particular distribution that describes brandsize choice probabilities across the population for each household's first purchase. However, after the first purchase, future brandsize choice probabilities are described by different distributions for each household. That is, each household has its own set of parameters, which are updated after each household's purchase occasion by Bayesian methods.

Such an approach makes use of properties of the conjugate distributions, Dirichlet and multinomial. This pair has important advantages for sequential Bayesian sampling of a multinomial process (frequently assumed for describing product purchases). If the prior distribution of probabilities is Dirichlet, then the posterior is too, and its parameters are a simple update of the prior parameters with the sample information. Therefore, given a sequence of purchases by a customer, the Dirichlet-multinomial makes it possible to learn the underlying vector of purchase probabilities for the household with greater and greater precision, provided, of course, that the assumptions of the model are satisfied. We shall describe this Bayesian update view of the Dirichlet-multinomial as the *dynamic* use of the Dirichlet.

The Dirichlet, as a *static* distribution, has been widely exploited in studying purchase behavior by Chatfield, Ehrenberg and Goodhardt (1966), Goodhardt, Ehrenberg and Chatfield (1984), and Jeuland, Bass and Wright (1980) among others. These authors assume stationarity, in fact, zero-order conditions, and so do not explicitly consider marketing activities. Yet, marketing managers are paid to make markets non-stationary in favor of their own brands. And, indeed, it is more appealing and, with scanner panel data possible, to bring marketing variables into the model.

Fader (1993) and Fader and Lattin (1993) introduce marketing variables into the dynamic Dirichlet-multinomial model. Fader (1993) constructs a composite of the Dirichlet process for discovering household loyalty with a multinomial logit for modeling the effects of marketing on household choice. In Fader and Lattin (1993) households change their whole purchase loyalty vector at random times picked by a recurrent event process, whereupon the Dirichlet update starts over again. Neither of these papers take up filtering or purchase feedback.

A number of authors have introduced explanatory variables into Dirichlet parameters. Heckman and Willis (1977) do so in a study of labor force participation. They model the binary or beta-logistic case but note that the approach extends naturally to the multivariate Dirichlet formulation. Dunn and Wrigley (1985), in a study of urban shopping behavior, introduce explanatory variables into the Dirichlet parameters in the manner suggested by Heckman and Willis and which we shall employ here. The approach leads naturally to a multinomial logit. Their application does not need to be concerned with filtering and purchase feedback. Uncles (1987) uses the beta-logistic to introduce explanatory variables into choice of mode of travel on shopping trips. Davies (1984) models time-varying and feedback effects in residential mobility in a binary case, using what he terms the generalized beta-logistic. Filtering is not an issue in either of these papers.

Our model draws on previous research but seeks to devise and calibrate a model that embraces all three phenomena: that is, we would like not only to infer underlying preference, but also to measure and filter out marketing effects and to introduce purchase feedback. An overall view of the model is as follows: purchase probabilities for a household are assumed to have a prior distribution that is Dirichlet before purchase. The parameters of the distribution, however, are functions of marketing and possibly other explanatory variables. Purchase then has a multinomial distribution with sample size one. The posterior distribution is determined and, from its parameters, a more precise view of household preference at the time of purchase is determined. As this is being done, marketing effects are filtered out. Experience with the product is assumed to affect preferences for subsequent purchases according to a simple learning model, thereby introducing purchase feedback. At this point we consider the possibility that unmeasured disturbances may also affect product choice and so we build in a tendency for choice to revert to market norms. Out of these several processes and actions come a new set of Dirichlet parameters with which to start the next purchase cycle.

The paper develops the model in the following steps: prior distribution, marketing model, sampling, posterior distribution, purchase feedback, other disturbances, and next cycle. Then the model is calibrated on two different databases and compared to other models in the literature. Finally, since purchase feedback implies that current marketing creates future value, we seek insight about the amount and duration of such effects. This is done by developing recursive expressions for expected sales and applying them to determine analytic formulas for the future value of current marketing actions under special scenarios.

## MODEL DEVELOPMENT

### Prior Distribution

To explain the model formulation, we focus on one household. (At calibration we shall work with all.)

- B = the number of products being considered;
- $\phi$  =  $[\phi_1, \dots, \phi_B]$   
= a vector of purchase probabilities for the products,  $k = 1, \dots, B$ ;
- t = index of purchase occasions,  $t = 1, 2, \dots$ ;
- $\mathbf{x}_{kt}$  = a column vector of marketing and other explanatory variables for product k at occasion t; (for brevity we shall frequently refer to  $\mathbf{x}_{kt}$  simply as "marketing.")

$$\mathbf{x}_t = [x_{1t}, \dots, x_{Bt}]$$

$f_t(\phi | \mathbf{x}_t)$  = prior distribution of  $\phi$  at  $t$ , given  $\mathbf{x}_t$ .

We take the prior to be Dirichlet:

$$f(\phi | \mathbf{x}) = \frac{\Gamma(s)}{\Gamma(\alpha_1) \dots \Gamma(\alpha_B)} \phi_1^{\alpha_1-1} \dots \phi_B^{\alpha_B-1} \quad (1)$$

where  $\alpha_k \equiv \alpha_k(\mathbf{x})$ ,  $s \equiv s(\mathbf{x}) = \sum_k \alpha_k(\mathbf{x})$ , and  $\sum_k \phi_k = 1$ . In the absence of other information we take the prior distribution for a household to be based on the cross sectional distribution of purchase probability vectors for the population in question. The underlying viewpoint is that household choices are heterogeneous as a result of many unobserved and unobservable influences (Davies, 1984). A randomly chosen household has a prior Dirichlet distribution with the parameters of the population. Subsequently, as a result of observing the household's purchases, we resolve some, but not all, of the uncertainty associated with the original heterogeneity.

The mean and covariance of  $\phi$  are:

$$\bar{\phi}_k(\mathbf{x}) = \frac{\alpha_k(\mathbf{x})}{s(\mathbf{x})} \quad (2)$$

$$\text{cov}(\phi_j, \phi_k) = \frac{1}{s(\mathbf{x})+1} \frac{\alpha_j(\mathbf{x})}{s(\mathbf{x})} \left[ \delta_{jk} - \frac{\alpha_k(\mathbf{x})}{s(\mathbf{x})} \right] \quad (3)$$

where  $\delta_{jk}=1$ , if  $j=k$ , otherwise 0 (the Kronecker delta). A particularly useful property of the Dirichlet is that the marginal densities for each of the variables are beta:

$$f(\phi_k | \mathbf{x}) = \frac{\Gamma(s)}{\Gamma(\alpha_k) \Gamma(s-\alpha_k)} \phi_k^{\alpha_k-1} (1-\phi_k)^{(s-\alpha_k)-1} \quad (4)$$

Most quantities depend on  $t$ , the purchase occasion, but we suppress its subscript unless needed for clarity.

The Dirichlet has several interesting features for our application. First, although it provides a joint density of  $B$  random variables, each with its own mean, it has only  $B$  independent parameters. The means are constrained to be non-negative and sum to one, leaving one more parameter, which, in conjunction with the means, serves to determine all  $B(B+1)/2$  covariances. This is quite parsimonious. It also implies, as may be seen from (3), that increasing  $s$ , while holding all the means constant, decreases all the covariances together. Thus, as inspection of (2) and (3) shows, the means and variances of the Dirichlet are quite tightly tied

together. Obviously more general distributions could be considered. We choose the Dirichlet because (1) it has a history of successful use on products such as those we consider, (2) it is parsimonious and (3) its analytic tractability leads to results that are relatively easy to interpret.

## Marketing Model

The effects of marketing and other explanatory variables will enter the model through the Dirichlet parameters,  $\alpha_k$ . In this section we argue for a particular formulation and show that it leads naturally to the familiar multinomial logit model for the expected probability of product choice.

The model for  $\alpha$  is:

$$\alpha_{kt}(x_t) = \alpha_{kt}(0) e^{\beta x_{kt}} \quad (5)$$

where  $\beta$  = a row vector of coefficients and  $\alpha_k(0)$  is a constant.

The use of an exponential function to express the effect of explanatory variables on  $\alpha$  in the Dirichlet is fairly common. Heckman and Willis (1977) and Dunn and Wrigley (1985), employ it, as do Davies (1984) and Uncles (1987).

Since  $\beta x_{kt}$  is linear,  $\alpha_k$  is actually multiplicative in separate functions of each variable. From a marketing point of view, if a price reduction creates a 20% increase in, say, a price component of  $\alpha_k$ , and a special display creates a 20% increase in a display component, then the model produces a net effect on  $\alpha_k$  of  $1.2 \times 1.2 = 1.44$ , or 44%. This type of built-in multiplicative interaction among the effects of explanatory variables seems appropriate for many marketing activities. It is similar to that used in a general marketing mix model by Little (1975).

A methodological point about our formulation is that we do not apply the marketing force,  $e^{\beta x}$ , directly to each household's (unobservable) purchase probability vector,  $\phi$ , but rather to the parameters of their distribution, i.e.  $\alpha_k$ . As indicated above, this has frequently been done. The formulation has an interesting advantage. If we consider a set of households characterized by the Dirichlet parameters,  $\{\alpha_k(x)\}$ , the model does not require that marketing affect each of these households in the same amount and direction. Response, in other words, may be heterogeneous. We only postulate that the means move as determined by (5) and, correspondingly, that the rest of the distribution is Dirichlet with the parameters (5).

For purposes of exposition, we are considering only one household. However, the parameter set,  $\{\alpha_k\}$ , is updated by household, based on marketing variables and actual brand choice. Thus, in full generality, there is a household subscript on  $\alpha_k$ . Prior to the first



purchase occasion in the dataset for a given household, we have no information and so assign each household the same set of parameters  $\{\alpha_{k,1}(0)\}$ . After the first purchase occasion, each household has its own parameters that will usually be unique to the household.

As seen in (5), we have chosen  $\mathbf{x}=0$  as the reference vector for the marketing model. Doing so loses no generality, since we can pick any other reference vector and measure the explanatory variables from it. In this way a reference can depend on the product and even the purchase occasion. We shall frequently do this in our applications. For example, the price of a product in a given store is often measured relative to a base year average for that store. The calibration of the model in relation to reference values is a useful device in practice (Little, 1975).

As already noted, the actual probability of purchase,  $\phi_k$ , is unobservable. In order to estimate the parameters of its distribution by maximum likelihood using the observed data, we need the expected value of  $\phi_k$  as a function of the parameters. This is

$$p_k(\mathbf{x}) = E\{\phi_k|\mathbf{x}\} = \bar{\phi}_k(\mathbf{x}) \quad (6)$$

or, using the marketing model,

$$p_k(\mathbf{x}) = \frac{\alpha_k(0)e^{\beta x_k}}{\sum_j \alpha_j(0)e^{\beta x_j}} \quad (7)$$

Another way of writing this is

$$p_k(\mathbf{x}_t) = \frac{e^{v_k(\mathbf{x}_t)}}{\sum_j e^{v_j(\mathbf{x}_t)}} \quad (8)$$

where

$$v_k(\mathbf{x}_t) = \ln \alpha_{kt}(0) + \beta x_{kt} \quad (9)$$

and the dependence on  $t$  is explicit.

The expected choice probability (8) is the multinomial logit. The quantity,  $\ln \alpha_{kt}(0)$ , plays a similar role to that played by loyalty in the Guadagni-Little model but is hard-wired in the log form with unity coefficient. Notice that, instead of  $\ln \alpha_k(0)$ , we can equivalently use  $\ln \bar{\phi}_k(0)$ , where

$$\bar{\phi}_k(0) = \frac{\alpha_k(0)}{\sum_j \alpha_j(0)} \quad (10)$$

The reason is that the denominator of (10) would cancel out of (8). When (9) is expressed in this form, we see that loyalty is replaced by the log of the probability that the household will purchase the product under reference market conditions.

## Sampling

The product actually purchased is assumed to be the result of drawing a sample of size one from a multinomial distribution with probability vector,  $\phi$ . Let

$$\begin{aligned} r_k &= \text{the number of units of product } k \text{ drawn;} \\ \mathbf{r} &= [r_1, \dots, r_B]; \\ r &= \sum_k r_k = \text{sample size.} \end{aligned}$$

Then

$$P(\mathbf{r}|\phi) = \frac{r!}{r_1! \dots r_B!} \phi_1^{r_1} \dots \phi_B^{r_B} \quad (11)$$

In our case  $r=1$  and so the coefficient of factorials becomes unity.

## Posterior Distribution

The product that is bought on purchase occasion  $t$  gives us new information about the joint distribution of the purchase probabilities,  $\phi$ . The posterior distribution of  $\phi$  is determined by Bayes rule:

$$f^{\text{post}}(\phi|\mathbf{r}_t, x_t) = \frac{P(r_t|\phi) f(\phi|x_t)}{\int P(r_t|\phi) f(\phi|x_t) d\phi} \quad (12)$$

From Dirichlet-multinomial theory, the posterior is Dirichlet with

$$\alpha_{kt}^{\text{post}}(x_t) = \alpha_{kt}(x_t) + \mathbf{r}_{kt} \quad (13)$$

but, using the marketing model (4),

$$\alpha_{kt}^{\text{post}}(x_t) = \alpha_{kt}^{\text{post}}(0) e^{\beta x_{kt}} = \alpha_{kt}(0) e^{\beta x_{kt} + \mathbf{r}_{kt}}. \quad (14)$$

Therefore,

$$\alpha_{kt}^{\text{post}}(0) = \alpha_{kt}(0) + r_{kt} e^{-\beta x_{kt}} . \quad (15)$$

This is the standard Dirichlet update, except that the marketing model has introduced an adjustment, which, following Lattin (1991), we call *filtering*.

Filtering addresses a problem that has bothered many people. If a product was bought on promotion, should not that purchase exert less influence in updating a loyalty measure? Equation (14) takes care of this. Suppose, for example, that a special display for product  $k$  at  $t$  creates a large value for  $\beta x_{kt}$ . This makes it more likely that  $k$  will be bought. Therefore, the update formula reduces the weight given to  $r_{kt}$  relative to  $\beta x_{kt}=0$  in the calculation of  $\alpha_{kt}^{\text{post}}(0)$ . For the same reason, less precision is added to the updated Dirichlet distribution of  $\phi$ . In this way marketing effects are removed from our loyalties.

It may be helpful to relate our formulation at this point to other models in the literature. For  $\beta=0$  (no marketing in the model), we have a Dirichlet-multinomial model in the form used by Fader (1993). If we keep  $\beta x_{kt}$ , but do not perform the filtering, we have the Fader (1993) DM-MNL model.

To complete the posterior formulas:

$$s_t^{\text{post}}(0) = s_t(0) + \sum_k r_{kt} e^{-\beta x_{kt}} \quad (16)$$

$$\bar{\phi}_{kt}^{\text{post}}(x) = \frac{\alpha_{kt}^{\text{post}}(0) e^{\beta x_k}}{\sum_j \alpha_{jt}^{\text{post}}(0) e^{\beta x_j}} \quad (17)$$

where we have written  $x$  without a  $t$  subscript to emphasize that this is a functional relation good for any  $x$ . Specifically for  $x = 0$ ,

$$\bar{\phi}_{kt}^{\text{post}}(0) = \frac{\alpha_{kt}^{\text{post}}(0)}{s_t^{\text{post}}(0)} . \quad (18)$$

## Purchase Feedback

The idea behind purchase feedback is that experience with a product will tend to increase the probability of its repurchase in the future. Of course, some people may have a poor experience with a product, but unless there is a net positive experience in the population, the item will presumably disappear from the market. We hypothesize the existence of feedback, and model it with a parameter,  $\gamma$ , that can range from no feedback ( $\gamma=1$ ) to complete forgetting of

the past ( $\gamma=0$ ). Furthermore, we do not require that feedback has to happen the same way for each household and each purchase occasion or even always be positive. This is because the Dirichlet permits heterogeneity of probabilities. Feedback, however, is presumed to shift the *mean* in favor of the product purchased. Specifically we suppose:

$$\bar{\phi}_{kt}^{\text{feed}}(0) = \gamma \bar{\phi}_{kt}^{\text{post}}(0) + (1 - \gamma)r_{kt} \quad (19)$$

where  $\bar{\phi}_{kt}^{\text{feed}}(0)$  = the household's expected purchase probability for k under reference marketing conditions after experience with the product purchased at occasion t.

It would be possible to introduce multiple  $\gamma$ 's, representing different feedback conditions. For example, some writers have suggested that products bought on promotion have less influence on future purchases than those that are not. This could be incorporated. However, prudence suggests that we start simply with a single  $\gamma$ .

Notice that  $\gamma$  speaks to one of the main goals of our paper. Since we have previously used  $r_{kt}$  in the Dirichlet update of  $\bar{\phi}_{kt}^{\text{post}}(0)$ , we are using the observed purchase,  $r_{kt}$ , twice. This implements our separation of past product preference of the household (determined better by the Dirichlet update) and purchase feedback (represented by the learning model (19)). Therefore, the estimation of  $\gamma$  should help answer the question of whether the historical success of loyalty-like variables is entirely due to their explaining heterogeneity and there really is no purchase feedback. This would be the case if we find  $\gamma=1$ .

Next we need to convert the feedback effects modeled by (19) into expressions for the  $\alpha$ 's and  $s$ . We suppose that the feedback process itself does not affect  $s_t^{\text{feed}}$ , which controls the variance of  $\phi$ . This is because  $r_{kt}$  is known with certainty at this stage of the process. Then

$$s_t^{\text{feed}}(0) = s_t^{\text{post}}(0) , \quad (20)$$

Defining  $\alpha_{kt}^{\text{feed}}(0)$  by

$$\bar{\phi}_{kt}^{\text{feed}}(0) = \frac{\alpha_{kt}^{\text{feed}}(0)}{s_t^{\text{feed}}(0)} , \quad (21)$$

a little algebra applied to (19), (20) and (21) yields:

$$\alpha_{kt}^{\text{feed}}(0) = \gamma \alpha_{kt}^{\text{post}}(0) + (1 - \gamma)r_{kt} s_t^{\text{post}}(0) . \quad (22)$$

We note that, as in the case of the marketing model, the feedback model does not work on the unobservable  $\phi$  but on the parameters of its Dirichlet distribution. Again this has the advantage that every household in the same state (i.e., having same set of  $\alpha_{kt}$ ) does not need to respond in the same way.

## Other disturbances

Next consider the effect of other possible disturbances on the purchase probabilities. At each purchase occasion the Dirichlet update increases the precision with which we know the household's underlying preference vector,  $\phi$ , because  $s_t(0)$  is incremented with a positive number. Perhaps this is whole story. If so,  $s_t(0)$  will increase with each purchase and, eventually, given a long enough time and large enough  $t$ , we shall know each household's preferences,  $\phi$ , almost perfectly. From a practical point of view we doubt that this happens. It seems likely that new, unobserved disturbances will introduce uncertainty into the purchase probabilities.

This idea has been introduced by Fader and Lattin (1993). They assume that at certain, randomly chosen, purchase occasions a household forgets its purchase history and makes a new draw of a purchase probability vector from the Dirichlet distribution representing the underlying population of all households. The probability of this happening is determined by a recurrent event process that has a geometric distribution of times between events. The parameter of this distribution is estimated simultaneously with the rest of the model.

Our approach is in the same spirit but operationally quite different. We suppose that unobserved influences in the market tend to bring a household back toward market norms. The disturbances will alter the purchase probability vector and make it look more as if it were drawn from the Dirichlet distribution for the whole population. We model this with a pair of processes that mix current means and variances with underlying population values.

Before proceeding, it will be helpful to define a special notation for the Dirichlet parameters at the start of the process. Let

$\tau_k \equiv \alpha_{k,1}(0)$  = the household's Dirichlet parameter for product  $k$  under reference marketing conditions on the first purchase occasion ( $t=1$ ).

$$s(0) = \sum \tau_k ,$$

$$\bar{\phi}^0 = \tau/s(0) ,$$

where  $\tau = (\tau_1, \dots, \tau_B)$  and  $\bar{\phi}^0$  is, therefore, the initial vector of mean probabilities under reference marketing. The model uses the same  $\tau$  for all households, i.e., each household's starting purchase probabilities at reference marketing are described by a Dirichlet distribution with parameters,  $\tau$ , and mean probabilities,  $\bar{\phi}^0$ . Every household's purchase probabilities at reference marketing is a draw from the same Dirichlet, which therefore may be described as the distribution for the population.

The mixing processes to represent the tendency for household means and variances to move toward population values are:

$$\bar{\phi}_{kt}^{\text{final}}(0) = \delta \bar{\phi}_{kt}^{\text{feed}}(0) + (1 - \delta)(\bar{\phi}_k^0), \quad 0 \leq \delta \leq 1. \quad (23)$$

$$\frac{1}{s_t^{\text{final}}(0)} = \theta \frac{1}{s_t^{\text{feed}}(0)} + (1 - \theta) \frac{1}{s(0)}, \quad 0 \leq \theta \leq 1. \quad (24)$$

where  $\delta$  and  $\theta$  are new smoothing constants. Notice that the calibration is free to choose  $\delta$  or  $\theta$  close to 1, which would imply little unobserved influence. Then the measurement of a household's preferences would become better and better up to a high degree of precision as purchase occasions accumulate. It is because we think that this will not be true for long time series that we include the mixing with population values.

Other interesting models of moving toward market norms would also be possible. Often the whole market changes as new products are introduced and customer habits evolve. For long time series, it might then be desirable, for example, to characterize the market norm as a moving average of previous shares.

To complete the description of processes during  $t$ , we define  $\alpha_{kt}^{\text{final}}(0)$  by means of

$$\bar{\phi}_{kt}^{\text{final}}(0) = \frac{\alpha_{kt}^{\text{final}}(0)}{s_t^{\text{final}}(0)}. \quad (25)$$

### Next Cycle

The cycle for purchase occasion  $t+1$  starts with

$$\begin{aligned} \alpha_{k,t+1}(0) &= \alpha_{kt}^{\text{final}}(0) \\ s_{t+1}(0) &= s_t^{\text{final}}(0). \end{aligned} \quad (26)$$

After substitutions, these becomes

$$\alpha_{k,t+1}(0) = s_{t+1}(0) \{ \delta (1 - \gamma) r_{kt} + (1 - \delta) \bar{\phi}_k^0 + \delta \gamma \alpha_{kt}^{\text{post}}(0) / s_t^{\text{post}}(0) \} \quad (27)$$

$$s_{t+1}(0) = s(0)s_t^{\text{post}}(0)/[(\theta)s(0) + (1-\theta)s_t^{\text{post}}(0)] \quad (28)$$

where

$$\alpha_{kt}^{\text{post}}(0) = \alpha_{kt}(0) + \tau_{kt} e^{-\beta x_{kt}}. \quad (29)$$

$$s_t^{\text{post}}(0) = s_t(0) + \sum_k \tau_{kt} e^{-\beta x_{kt}} \quad (30)$$

The parameters for purchase occasion  $t+1$  have now been expressed recursively in terms of their values at  $t$ , the observed data at  $t$ , and the  $t$ -invariant parameters,  $\beta$ ,  $\gamma$ ,  $\delta$ ,  $\theta$  and  $\tau$ .

### Initialization

The purchase process for a household starts at  $t=1$ . The purchase probabilities at reference marketing for each household have a Dirichlet distribution with parameters  $\tau$ . Thus

$$\alpha_{k1}(0) \equiv \tau_k$$

Since each household has this distribution, it is also the distribution across the population for  $t=1$ , and, as such, characterizes the heterogeneity of underlying preferences in the population.

## CALIBRATION

### Estimation of Parameters

The unknown parameters of the model are the population heterogeneity vector,  $\tau$ , the vector of coefficients of marketing and other explanatory variables,  $\beta$ , the purchase feedback constant,  $\gamma$ , and the parameters to introduce other disturbances,  $\delta$  and  $\theta$ . The recursive relations (27) and (28) express  $\alpha_{kt}^h(0)$  for each household,  $h$ , in terms of previous values, the unknown parameters, and the data. Expressions (8) and (9), with the addition of a household subscript, convert the estimation problem into multinomial logit form. The expressions are highly nonlinear, but the estimation procedure of Fader, Lattin, and Little (1992) determines maximum likelihood estimates of the parameters quite efficiently.

We calibrate our model and three others on two IRI panel datasets, one for juice drinks, the other for laundry detergents. Both come from the Eau Clair, Wisconsin, market. In each case we model the top ten brandsizes as measured by unit share. A random sample of households is drawn from each dataset. The time spans 39 weeks for juices and 52 for

detergents. The juice database contains 182 households and 1054 purchases in the calibration sample and another 46 households and 282 purchases in a holdout sample. The detergent database has 594 households and 2776 purchases for calibration plus 164 households and 739 purchases for holdout. The juice database was screened to include only households that made less than thirty and more than 1 purchase over the 39 weeks. Only a few households were thereby eliminated.

The marketing variables included in the model are display, feature and price. Our model assumes that every marketing variable has a reference value. Display and feature, which are each 0-1 variables, are given a reference value of 0. Price is defined as normalized price per unit, where the unit is either ounces for juices or washloads for detergents and the normalization is the average price per unit over a 26 week period in the store where the purchase is being made. Thus the price variable represents a percent deviation from the average price in the store. Clearly alternative assumptions and specifications of reference points are possible for any of the marketing variables.

Each database contains the top 10 brandsizes by unit volume. The names and market shares over the calibration period follow:

<b><u>Juice Database</u></b>	
<b>Brandsize</b>	<b>Market Share</b>
Ocean Spray Cranberry Juice 48oz	0.16
Gatorade Fruit Punch 32oz	0.09
Hawaiian Punch Red Drink 46oz	0.10
Gatorade Orange 32oz	0.13
Ocean Spray CranRaspberry Juice 48oz	0.08
Gatorade Lemon Lime 32oz	0.10
Ocean Spray Cranberry Juice 48oz	0.10
Ocean Spray Cranberry Juice 64oz	0.09
Ocean Spray CranRaspberry Juice 64oz	0.08

Two Ocean Spray Cranberry Juice 48oz products have been listed and are treated separately because their UPCs, prices, and merchandising activities differ.

<b><u>Detergent Database</u></b>	
<b>Brandsize</b>	<b>Market Share</b>
Break Thru Liquid 126oz	0.11
Tide Liquid 128oz	0.10
Solo Liquid 128oz	0.07
Tide Liquid 96oz	0.08
Ultra Cheer Powder 42oz	0.11
Ultra Tide Powder 42oz	0.10



Surf Liquid 64oz	0.13
YES Liquid 64oz	0.14
ERA+ Liquid 128oz	0.05
Oxydol Ultra Powder 42oz	0.11

## Results

Several comparisons are useful. First, phenomena are added to the model one at a time to show their incremental effects. Then the final model is compared to two previous models, Guadagni and Little (1983), which it is designed to improve, and Fader and Lattin (1993) which has been found to be quite good. The models are further tested on holdout samples.

Adding phenomena one at a time creates six nested models. Each may be thought of as a multinomial logit models but with its own version of utility. The first two correspond to existing models in the literature.

**(M1) Pure dynamic Dirichlet (Fader, 1993).** A Dirichlet model without marketing variables results from forcing  $\beta=0$  and setting  $\gamma=\delta=1$  in the new model.

$$\begin{aligned} \text{Model:} & \quad v_{kt} = \ln(\alpha_{kt}) \\ \text{Update:} & \quad \alpha_{k,t+1} = \alpha_{kt} + r_{kt} \end{aligned}$$

**(M2) Add marketing, Fader DM-MNL (Fader, 1993).** Fader's DM-MNL is one of the simplest way to add marketing to the Dirichlet and results from dropping filtering and setting  $\gamma=\delta=1$ .

$$\begin{aligned} \text{Model:} & \quad v_{kt} = \ln(\alpha_{kt}) + \beta \mathbf{x}_{kt} \\ \text{Update:} & \quad \alpha_{k,t+1} = \alpha_{kt} + r_{kt} \end{aligned}$$

**(M3) Add filtering.** M3 is obtained by setting  $\gamma=\delta=\theta=1$  in the full model and will show the effect of adding filtering to Fader's DM-MNL.

$$\begin{aligned} \text{Model:} & \quad v_{kt} = \ln(\alpha_{kt}) + \beta \mathbf{x}_{kt} \\ \text{Update:} & \quad \alpha_{k,t+1} = \alpha_{kt} + r_{kt} \exp(-\beta \mathbf{x}_{kt}) \end{aligned}$$

**(M4) Add purchase feedback.** M4 results from setting  $\delta=\theta=1$ .

$$\begin{aligned} \text{Model:} & \quad v_{kt} = \ln(\alpha_{kt}) + \beta \mathbf{x}_{kt} \\ \text{Update:} & \quad \alpha_{k,t+1} = \gamma \alpha_{kt} + r_{kt} ((1-\gamma)s_t + \exp(-\beta \mathbf{x}_{kt})) \end{aligned}$$

**(M5) Add return toward population mean.** Setting  $\theta=1$  produces M5.

$$\begin{aligned} \text{Model:} \quad & v_{kt} = \ln(\alpha_{kt}) + \beta \mathbf{x}_{kt} \\ \text{Update:} \quad & \alpha_{k,t+1} = \delta \{ \gamma \alpha_{kt} + (1-\gamma) r_{kt} [s_t + \exp(-\beta \mathbf{x}_{kt})] \} \\ & \quad + (1-\delta) \bar{\alpha}_k^0 [s_t + \sum_j r_{jt} \exp(-\beta \mathbf{x}_{jt})] \end{aligned}$$

**(M6) Add return toward population variance.** M6 is the full model.

$$\begin{aligned} \text{Model:} \quad & v_{kt} = \ln(\alpha_{kt}) + \beta \mathbf{x}_{kt} \\ \text{Update:} \quad & \text{See Equation (27)} \end{aligned}$$

Tables 1 and 2 show the results for the six models on the two databases. The measures of fit, namely, loglikelihood and adjusted  $\rho^2$ , generally improve as phenomena are added to the model. Adding marketing variables to the pure dynamic Dirichlet produces a dramatic jump in fit. Filtering brings about a small improvement. Purchase feedback results in another jump. The final two steps of adding a tendency to return to population norms produce small increases except for adding the last parameter in juices, which leaves the loglikelihood unchanged.

It is interesting and rather comforting to note that the coefficients for the marketing variables do not change much across the runs. All are within about a standard error of each other. In the final model the value of  $(1-\gamma)=0.12$  for juices (0.12 for detergents) demonstrates statistically significant purchase feedback. The value of  $(1-\delta)=0.02$  for juices (0.03 for detergents) supports the hypothesis of a tendency to revert to population means and  $(1-\theta)=0.04$  for juices (0.13 for detergents) suggests that unobserved disturbances limit how accurately we can learn individual household preferences.

As one might expect, the  $\tau$  parameters of the Dirichlet for the initial product preferences at reference marketing are approximately proportional to brandsize market shares.

### Comparisons with other models

Tables 3 and 4 show that the new model fits the data better than the models of Guadagni and Little (1983) and Fader and Lattin (1993). To be fair, we note that the comparison handicaps Guadagni and Little relative to their original paper (and common practice) by not providing a pre-period to initialize loyalty. The longer the preperiod, the more the loyalties will have settled down and so the more they will account for heterogeneity, right from the beginning of the calibration sample. This will increase  $\rho^2$ . On the other hand, the comparison here gives each model the same amount of data. Under the present circumstances the new model is better than either of the others, not only for the calibration period sample of 1054 purchases for juices (2776 for detergents), but also for the holdout sample of 282 purchases for juices (739 for detergents). The holdout samples come from different households in the same time periods.

**Table 1.** Juice database. Adding the features of the model one at a time shows progressively better fits without much change in the coefficients for the marketing variables.

<b>Model Coefficients (t-statistics in parentheses)</b>												
	(1) Pure Dirichlet		(2) Add Marketing		(3) Add Filtering		(4) Add Feedback ( $\gamma$ )		(5) Add Return to Market Mean ( $\delta$ )		(6) Add Return to Market Variance ( $\theta$ )	
<b>Price</b>			-4.58	(-4.9)	-4.55	(-5.4)	-4.77	(-5.2)	-4.79	(-5.2)	-4.78	(-5.2)
<b>Display</b>			0.84	(6.1)	0.90	(6.8)	0.95	(6.8)	0.94	(6.6)	0.94	(6.7)
<b>Feature</b>			0.29	(1.3)	0.06	(0.3)	0.18	(0.8)	0.22	(1.0)	0.21	(0.9)
$\tau_1$	0.17	(6.5)	0.20	(6.4)	0.14	(6.3)	0.24	(4.5)	0.24	(4.0)	0.22	(3.7)
$\tau_2$	0.12	(5.6)	0.11	(5.4)	0.08	(5.2)	0.13	(4.1)	0.13	(3.7)	0.12	(3.5)
$\tau_3$	0.11	(5.6)	0.12	(5.3)	0.09	(5.4)	0.15	(4.1)	0.14	(3.7)	0.13	(3.5)
$\tau_4$	0.12	(5.9)	0.11	(5.6)	0.08	(5.6)	0.13	(4.3)	0.13	(3.8)	0.12	(3.6)
$\tau_5$	0.08	(4.9)	0.10	(4.8)	0.07	(4.8)	0.13	(3.9)	0.13	(3.5)	0.12	(3.3)
$\tau_6$	0.11	(5.7)	0.14	(5.6)	0.10	(5.5)	0.17	(4.2)	0.17	(3.8)	0.15	(3.5)
$\tau_7$	0.11	(5.5)	0.08	(5.2)	0.06	(5.1)	0.10	(4.0)	0.10	(3.7)	0.09	(3.5)
$\tau_8$	0.07	(4.7)	0.09	(4.6)	0.06	(4.6)	0.11	(3.8)	0.11	(3.4)	0.10	(3.3)
$\tau_9$	0.14	(6.0)	0.14	(5.8)	0.10	(5.8)	0.17	(4.3)	0.17	(3.9)	0.15	(3.6)
$\tau_{10}$	0.10	(5.0)	0.10	(4.9)	0.08	(4.9)	0.13	(3.9)	0.13	(3.5)	0.11	(3.4)
$s = \sum \tau_k$	1.12		1.19		0.86		1.45		1.44		1.31	
(1- $\gamma$ )							0.10	(4.4)	0.15	(4.9)	0.13	(3.2)
(1- $\delta$ )									0.02	(2.6)	0.02	(2.6)
(1- $\theta$ )											0.04	(0.7)
<b>Log likeli- hood</b>	-1500		-1405		-1402		-1387		-1381		-1381	
<b>No. of parameters</b>	10		13		13		14		15		16	
<b>Adjusted <math>\rho^2</math></b>	.378		.416		.417		.423		.425		.424	

Number of purchases = 1054

Number of households = 182

Number of brandsizes = 10

**Table 2.** Detergent database. The ranking of the models and general qualitative conclusions are same as for the juice database.

<b>Model Coefficients (t-statistics in parentheses)</b>												
	(1) Pure Dirichlet		(2) Add Marketing		(3) Add Filtering		(4) Add Feedback ( $\gamma$ )		(5) Add Return to Market Mean ( $\delta$ )		(6) Add Return to Market Variance ( $\theta$ )	
<b>Price</b>			-5.35	(-14.8)	-5.13	(-13.8)	-5.41	(-14.9)	-5.40	(-14.8)	-5.43	(-14.9)
<b>Display</b>			1.05	(12.6)	1.03	(11.7)	1.06	(12.9)	1.06	(12.9)	1.06	(13.0)
<b>Feature</b>			0.46	(16.8)	0.45	(16.7)	0.47	(17.1)	0.46	(17.1)	0.46	(17.0)
$\tau_1$	0.11	(9.4)	0.08	(8.2)	0.05	(7.8)	0.08	(6.5)	0.08	(6.1)	0.07	(6.1)
$\tau_2$	0.12	(10.0)	0.08	(8.5)	0.05	(7.7)	0.09	(6.7)	0.08	(6.3)	0.07	(6.2)
$\tau_3$	0.09	(9.2)	0.07	(8.3)	0.04	(7.5)	0.07	(6.5)	0.07	(6.1)	0.06	(6.0)
$\tau_4$	0.09	(8.9)	0.08	(8.1)	0.05	(7.7)	0.08	(6.5)	0.08	(6.1)	0.07	(6.1)
$\tau_5$	0.28	(10.8)	0.23	(9.6)	0.13	(8.8)	0.25	(7.1)	0.22	(6.6)	0.19	(6.5)
$\tau_6$	0.26	(10.7)	0.22	(9.3)	0.12	(8.6)	0.22	(7.0)	0.21	(6.5)	0.18	(6.4)
$\tau_7$	0.13	(9.0)	0.08	(7.3)	0.05	(7.1)	0.08	(6.0)	0.08	(5.7)	0.07	(5.6)
$\tau_8$	0.12	(9.5)	0.07	(7.5)	0.04	(7.2)	0.07	(6.2)	0.06	(5.9)	0.06	(5.9)
$\tau_9$	0.07	(8.2)	0.05	(7.5)	0.03	(6.5)	0.06	(6.0)	0.05	(5.7)	0.04	(5.6)
$\tau_{10}$	0.26	(10.6)	0.23	(9.2)	0.13	(8.4)	0.24	(6.8)	0.21	(6.4)	0.18	(6.3)
$s = \sum \tau_k$	1.53		1.21		0.69		1.23		1.13		0.97	
(1- $\gamma$ )							0.12	(8.4)	0.15	(8.5)	0.12	(6.2)
(1- $\delta$ )									0.02	(3.4)	0.03	(4.0)
(1- $\theta$ )											0.13	(2.0)
<b>Log likeli- hood</b>	-4124		-3051		-3033		-2975		-2968		-2960	
<b>No. of parameter s</b>	10		13		13		14		15		16	
<b>Adjusted <math>\rho^2</math></b>	0.353		0.521		0.523		0.532		0.533		0.534	

Number of purchases = 2776

Number of households = 594

Number of brandsizes = 10

Although improved fits are obviously desirable, it seems more important that the new model isolates the individual phenomena of marketing, loyalty and purchase feedback.

Turning to a key question that motivated the model, recall that we update a household's preferences twice, once with Bayes rule and then again with the purchase feedback step, thus separating two phenomena combined into one constant by Guadagni-Little loyalty. If Guadagni-Little loyalty were only a device to explain heterogeneity and no purchase feedback exists, then we would expect  $(1-\gamma)=0$ . Since  $(1-\gamma)$  is significantly positive with values 0.13 (juices) and 0.12 (detergents), it appears that this is not the case.

It is noteworthy that the parameters for  $\Sigma\tau_k$ ,  $(1-\gamma)$ , and  $(1-\delta)$  are rather similar for the two categories. The history of the Guadagni-Little model suggests a fairly confined range of  $(1-\gamma)$  across data sets and categories. Perhaps similar empirical generalizations will evolve for the new model.

## USING THE MODEL

So far we have focused on understanding the phenomena underlying household choice behavior. However, ultimately, the model should be put to practical use in evaluating marketing strategy and tactics. For this purpose we need to simulate and compare different marketing mix scenarios. To facilitate such calculations we develop recursive expressions to calculate expected sales. Then we examine a particularly interesting question prompted by the existence of purchase feedback: What is the future value of current marketing? Although the complete answer to such a question depends on the particular circumstances, we design special scenarios that offer analytic insight into the effects.

### Simulation of Expected Sales

To build the framework, we first note that the underlying model of choice probabilities for a single household at purchase occasion  $t$  is a Dirichlet distribution having vector of means  $\bar{\phi}_t(\mathbf{x})$  with a component for product  $k$  of

$$\bar{\phi}_{kt}(\mathbf{x}) = \frac{\bar{\phi}_{kt}(0)e^{\beta x_k}}{\sum_j \bar{\phi}_{jt}(0)e^{\beta x_j}} \quad (32)$$

This follows from (6), (7), and (10).

The dynamics of purchase feedback is given by (19). In vector form,

**Table 3.** Juice database. Relative to the models of Guadagni and Little (1983) and Fader and Lattin (1993), the new model fits the data somewhat better and, more important, separates the phenomena of marketing, loyalty, and feedback.

		Model Coefficients (t-statistics in parentheses)					
		(1) Guadagni- Little		(2) Fader-Lattin		(3) New Model	
<b>Price</b>		-4.51	(-4.8)	-4.38	(-4.7)	-4.78	(-5.2)
<b>Display</b>		0.96	(7.0)	0.84	(6.1)	0.94	(6.7)
<b>Feature</b>		0.18	(0.8)	0.33	(1.5)	0.21	(0.9)
<b>Brand Size</b>	<b>1</b>			0.15	(5.2)	0.22	(3.7)
	<b>2</b>	-0.47	(-2.8)	0.08	(4.7)	0.12	(3.5)
	<b>3</b>	-0.13	(-0.8)	0.09	(4.6)	0.13	(3.5)
	<b>4</b>	-0.20	(-1.3)	0.08	(4.8)	0.12	(3.6)
	<b>5</b>	-0.19	(-1.1)	0.08	(4.3)	0.12	(3.3)
	<b>6</b>	-0.14	(-0.9)	0.10	(4.7)	0.15	(3.5)
	<b>7</b>	-0.47	(-2.9)	0.06	(4.6)	0.09	(3.5)
	<b>8</b>	-0.31	(-1.8)	0.07	(4.2)	0.10	(3.3)
	<b>9</b>	-0.07	(-0.4)	0.10	(4.9)	0.15	(3.6)
	<b>10</b>	-0.17	(-1.0)	0.08	(4.3)	0.11	(3.4)
$\Sigma \tau_k$				0.90		1.31	
<b>Loyalty</b>		5.38	(19.1)				
<b>Feedback (1-<math>\gamma</math>)</b>		0.31	(2.4)			0.13	(3.2)
<b>Toward Market</b>				0.04	(2.7)	0.02	(2.6)
<b>Mean (1-<math>\delta</math>)</b>							
<b>Toward Market</b>						0.04	(0.7)
<b>Variance(1-<math>\theta</math>)</b>							
<b>Log likelihood</b>		-1435		-1402		-1381	
<b>No. of</b>		14		14		16	
<b>Parameters</b>							
<b>Adjusted <math>\rho^2</math></b>		.403		.417		.424	

Number of purchases = 1054

Number of households = 182

Number of brandsizes = 10

Hold out sample = 282 purchases, 46 Households

<b>Log likelihood</b>	-360	-341	-335
<b>Adjusted <math>\rho^2</math></b>	.425	.454	.460

**Table 4.** Detergent database. Again, the new model fits the data somewhat better than the older model and, more important, separates marketing, loyalty, and feedback.

**Model Coefficients (t-statistics in parentheses)**

		(1) Guadagni- Little		(2) Fader-Lattin		(3) New Model	
Price		-5.63	(-15.2)	-5.21	(-14.5)	-5.43	(-14.9)
Display		0.99	(12.0)	1.05	(12.6)	1.06	(13.0)
Feature		0.44	(16.4)	0.46	(16.8)	0.46	(17.0)
Brand Size	1			0.06	(7.0)	0.07	(6.1)
	2	-0.19	(-1.8)	0.06	(7.2)	0.07	(6.2)
	3	-0.18	(-1.6)	0.05	(7.0)	0.06	(6.0)
	4	-0.08	(-0.7)	0.06	(7.0)	0.07	(6.1)
	5	0.57	(5.3)	0.17	(7.6)	0.19	(6.5)
	6	0.65	(5.6)	0.16	(7.5)	0.18	(6.4)
	7	-0.13	(-1.1)	0.06	(6.5)	0.07	(5.6)
	8	-0.2	(-1.8)	0.05	(6.7)	0.06	(5.9)
	9	-0.36	(-2.9)	0.04	(6.4)	0.04	(5.6)
	10	0.79	(7.0)	0.17	(7.4)	0.18	(6.3)
$\Sigma \tau_k$				0.90		0.97	
Loyalty		5.48	(25.9)				
Feedback (1- $\gamma$ )		0.34	(3.5)			0.12	(6.2)
Toward Market				0.05	(5.0)	0.03	(4.0)
Mean (1- $\delta$ )							
Toward Market						0.13	(2.0)
Variance(1- $\theta$ )							
Log likelihood		-3065		-3043		-2960	
No. of		14		14		16	
Parameters							
Adjusted $\rho^2$		.518		.522		.534	

Number of purchases = 2776

Number of households = 594

Number of brandsizes = 10

Hold out sample = 739 purchases, 164 households

Log likelihood	-838	-828	-799
Adjusted $\rho^2$	.500	.505	.521

$$\bar{\phi}_t^{\text{feed}}(0) = \gamma \bar{\phi}_t^{\text{post}}(0) + (1 - \gamma) r_t \quad (33)$$

where  $\bar{\phi}_t^{\text{post}}(0)$  is the posterior estimate of  $\bar{\phi}_t(0)$  and  $r_t$  is the vector of the household's product choices at  $t$ . The choice  $r_t$  is a random variable with a multinomial distribution having unobservable actual probabilities,  $\phi$ , and expected prior probabilities  $p_t(x_t) = \bar{\phi}_t(x_t)$ . See (6). In addition to feedback the model includes a tendency for the household's choice to move toward the population preferences. This is accomplished through the mixing model (22) and parameter  $\delta$ . The resulting final vector of mean purchase probabilities at  $t$  initializes the next purchase occasion as follows:

$$\bar{\phi}_{t+1}(0) = \delta \gamma \bar{\phi}_t^{\text{post}}(0) + \delta (1 - \gamma) r_t + (1 - \delta) \bar{\phi}^0 \quad (34)$$

Note that the final step of mixing in population variance, (23), does not affect (34) because it does not change the means.

Viewing the process a priori, we calculate expected values at  $t$  before knowing the actual outcomes to obtain

$$E\{\bar{\phi}_t^{\text{post}}(0) | \bar{\phi}_t(0)\} = \bar{\phi}_t(0) \quad (35)$$

$$E\{r_t | \bar{\phi}_t(0)\} = p_t(x_t)$$

so that

$$E\{\bar{\phi}_{t+1}(0) | \bar{\phi}_t(0)\} = \delta \gamma \bar{\phi}_t(0) + \delta (1 - \gamma) p_t(x_t) + (1 - \delta) \bar{\phi}^0 \quad (36)$$

Conditioning the whole recursive process on the starting vector of mean choice probabilities,  $\bar{\phi}^0$ , we can appropriately use the notation  $\bar{\phi}_{t+1}(0)$  for the left hand side of (36). Then

$$\bar{\phi}_{t+1}(0) = \delta \gamma \bar{\phi}_t(0) + \delta (1 - \gamma) p_t(x_t) + (1 - \delta) \bar{\phi}^0 \quad (37)$$

Equation (37) describes the basic dynamics of the expected values of the household choice probabilities at reference marketing.

Therefore, (37) makes it quite easy to simulate the expected purchase of a household for any marketing scenario  $\{x_t\}$ . Starting from the population value,  $\bar{\phi}^0$ , (37) recursively calculates each  $\bar{\phi}_t(0)$ . Expected sales to the household at purchase occasion  $t$  are then calculated from



$p_t(\mathbf{x}_t) = \bar{\phi}_t(\mathbf{x}_t)$ , as given by (32).

## The Future Value of Current Marketing

An important goal of our work has been to isolate purchase feedback and understand its implications for practice. As discussed earlier, the underlying idea is that a product, once purchased, comes into the household and creates brand name exposure and product experience. These can be expected to increase the probability of future purchase. The implication is that current marketing generates future value.

In the course of developing the final model by adding phenomenon one at a time, we have defined a family of six models, each with somewhat different parameters and properties. We now examine several of these to see how their parameters affect the future value of current marketing. Proofs of the properties appear in an Appendix.

**Model M3:** This model contains no purchase feedback or return to population norms ( $\gamma=1$ ,  $\delta=1$ ). It has the following

*Property P1: Marketing at the current purchase occasion affects the household's choice only on that occasion.*

Therefore, in M3, marketing has no long term value. Although our estimation process is free to pick this case, it does not do so in our databases. We find  $(1-\gamma)=0.12$  (juice drinks) and  $(1-\gamma)=0.13$  (detergents), both of which indicate statistically significant and operationally important amounts of purchase feedback.

**Model M4:** This model has purchase feedback but no tendency to return to population preferences ( $\gamma < 1$ ,  $\delta=1$ ). It has

*Property P2: If all products use reference marketing ( $\mathbf{x}=0$ ) at all purchase occasions, except that product  $k$  does a special promotion at the current occasion, the purchase probability for  $k$  will increase now and will continue with a constant (but smaller) increase indefinitely into the future.*

Put another way, in M4, if competitors stick to reference marketing, a product's one-time marketing actions produce a permanent effect. In practice, most products have fairly regular promotional activities. This erodes the magnitude of the permanent effect and makes it hard to see in simulations that use historical marketing data. Nevertheless, marketing activities in M4 have effects that continue indefinitely into the future. P2 shows that this is an inherent property of the model.

Because a household's environment contains unobserved (at least by us) marketing, we have included a decay mechanism in our final model. Marketing managers work to gain market

share for their products and redouble their efforts if their products are losing share. This type of global feedback tends to stabilize shares. Therefore, we have let the decay be in the direction of population preferences, which are closely related to shares. For our two databases we find that the decay parameter is small but statistically significant with  $(1-\delta)=0.02$  (juice drinks) and 0.03 (detergents). The resulting final model M6 is the one best supported empirically in the two databases.

**Model M6:** The complete new model contains both purchase feedback and return toward population preferences ( $\gamma < 1$ ,  $\delta < 1$ ,  $\theta < 1$ ). It has the following

*Property P3:* Starting at purchase occasion  $t=s$ , consider the relative effectiveness of two marketing scenarios  $\{x_t\}$  and  $\{x_t^*\}$ , which differ at  $t=s$  but for  $t>s$  are both at reference marketing ( $x_t=x_t^*=0$ ). For a given household, let  $p_t = p_t(x_t)$  and  $p_t^* = p_t^*(x_t^*)$  and

$$\Delta p_t = p_t^* - p_t,$$

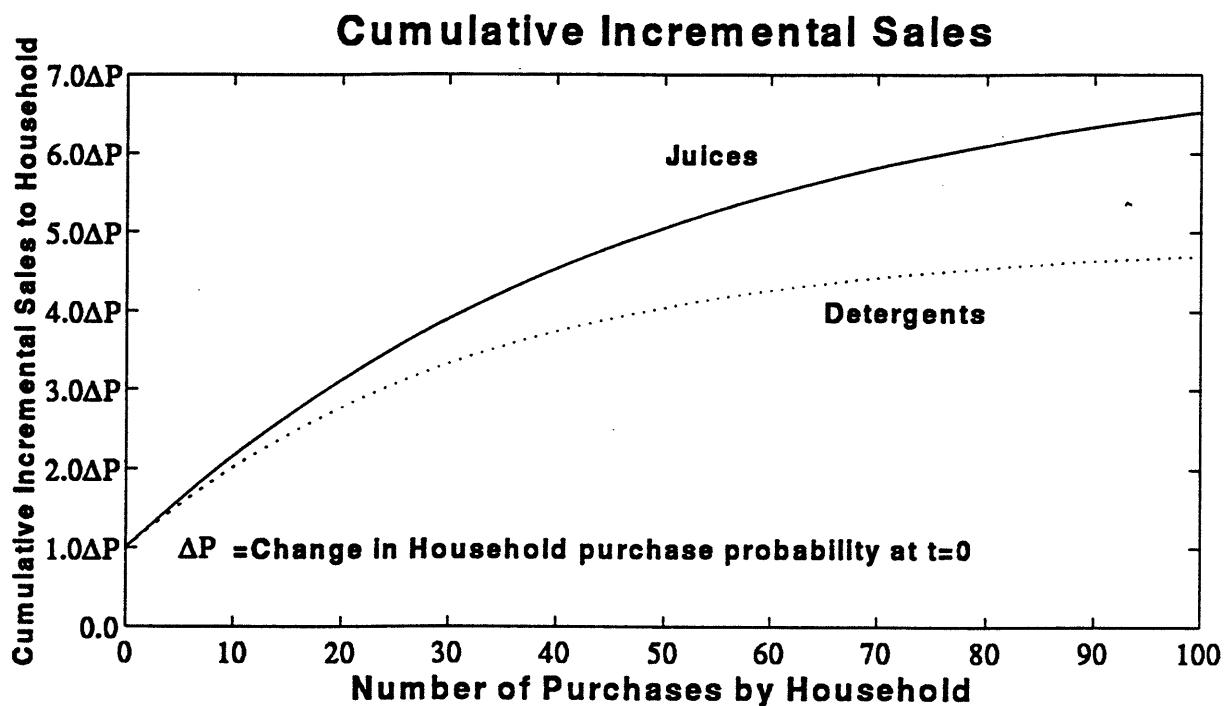
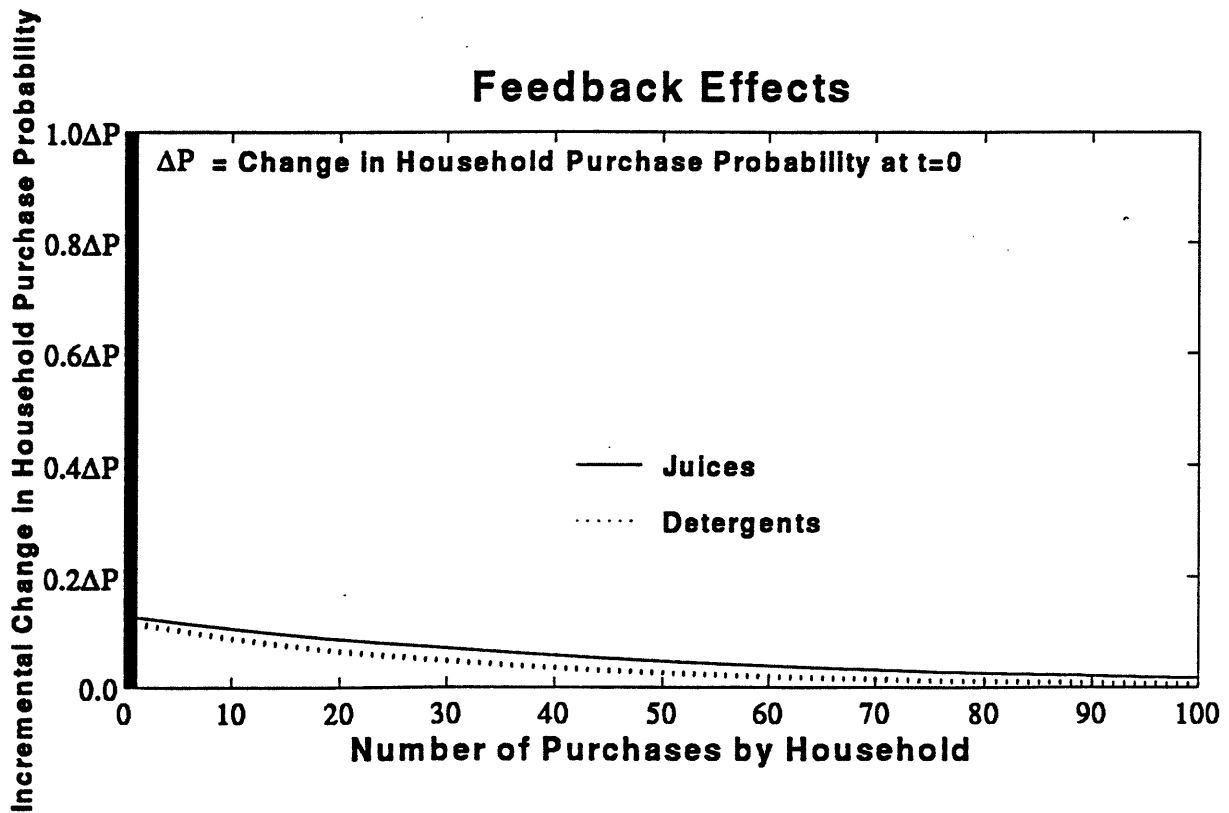
so that  $\Delta p_t$  is the difference in purchase probability vectors at  $t$  for the two scenarios. Then the incremental expected sales attributable to the  $\{x_t^*\}$  scenario over the  $\{x_t\}$  scenario for this household are:

- (a)  $\Delta p_s$  for  $t=s$ ,
- (b)  $\Delta p_t = (1-\gamma)\delta^{t-s} \Delta p_s$  for  $t>s$ , and
- (c)  $[(1-\gamma\delta)/(1-\delta)] \Delta p_s$  summed over all  $t \geq s$ .

Figure 1 shows the growth of total expected incremental sales of product  $k$  to the household as a result of a marketing event that produces a  $\Delta p_k$  increase in purchase probability at the first occasion and all other marketing activities hold at reference levels. We see that the accumulation of value over time from the marketing event is appreciable. The asymptote is 7.37  $\Delta p_k$  for juice drinks and 4.88  $\Delta p_k$  for detergents. It is interesting to note that, in the absence of competitive activities, the effect of product  $k$ 's marketing lasts quite long - for detergents it takes 17 purchase occasions to reach 50% of the total effect, for juice drinks, 28.

We need to realize that the scenario of property P3 will produce a relatively high incremental value, since all marketing is at reference levels except for the single pulse being evaluated. The reference marketing assumption does not hold if competitors are cutting price, adding displays, etc. Still, P3 highlights the value of purchase feedback. Simulation can determine the exact effect of more complex scenarios involving many households and actual historical marketing activities.

Property P3 also brings out another point. The incremental future value for product  $k$  is proportional to  $\Delta p_{kt}$ . This reminds us which customers are generating the most incremental



**Figure 1.** Future effects of a marketing event exceed current ones. The plot shows expected incremental sales to a particular household as a result of a single marketing event. The event produces an increase of  $\Delta p$  in purchase probability at the first purchase occasion. All other marketing activities are held at reference levels throughout. **Top:** incremental expected sales by purchase occasion for the product under consideration. **Bottom:** cumulative incremental expected sales.

sales.  $\Delta p_{kt}$  depends on the marketing action  $\mathbf{x}_{k,t}$  and also on the household's current probability of purchase. Applying marketing to a household with a prior preference close to either zero or 1 will not greatly shift that probability because of the nonlinear nature of logit probabilities. One can broadly classify households with extreme choice probabilities at reference marketing as loyal and other households as switchers. In model M6, when a marketing pulse is applied to a heterogeneous population, a loyal household's purchase probability will change a smaller amount than that of a switching household. Thus most of the incremental benefit of applying a marketing pulse comes from influencing the purchase probability of switching households.

## Conclusion

We have presented a model that partials out several key phenomena: loyalty, marketing response, and purchase feedback. The Dirichlet distribution and Bayesian update allow us to model household heterogeneity parsimoniously. From a methodological standpoint, the model tries to mimic customer and market processes. The filtering step addresses the issue that one learns less about a household's preference from a promotional purchase than a regular purchase because promotion confounds underlying preference and marketing influence. Separating purchase feedback from the loyalty update allows us to assess the value of experiencing a product. The positive value of the feedback parameter suggests that purchases do shift customer preferences toward the products purchased.

Given the presence of purchase feedback, current marketing has future value. We have developed benchmarks that help to provide insight about the amount and duration of such future value.

## APPENDIX: PROPERTIES OF THE MODEL

We here calculate certain properties of several models in the nested family of six that have been developed. Consider first M3 ( $\gamma=1, \delta=1$ ). Equation (37) becomes

$$\bar{\phi}_{t+1}(0) = \bar{\phi}_t(0) \quad . \quad (38)$$

Thus the household's mean purchase probabilities at reference marketing never change. The purchase probabilities at any purchase occasion are affected by the marketing at the time as shown in (32), but the effects do not carry forward. This establishes property P1.

For model M4 ( $\delta=1$ ), (37) becomes

Let  $t=s$  be the current purchase occasion. We see that, if  $\mathbf{x}_s \neq 0$  and  $\bar{\phi}_s(\mathbf{x}_s) \neq \bar{\phi}_s(0)$ , then

$$\bar{\phi}_{t+1}(0) = \gamma \bar{\phi}_t(0) + (1-\gamma) \bar{\phi}_t(\mathbf{x}_t) \quad . \quad (39)$$

$\bar{\phi}_{s+1}(0) \neq \bar{\phi}_s(0)$ . But, if thereafter,  $\mathbf{x}_t=0$  for  $t>s$ ,

$$\bar{\phi}_{t+1}(0) = \bar{\phi}_t(0) \quad t>s. \quad (40)$$

and so the vector of mean purchase probabilities at reference marketing is permanently changed. If  $\Delta \mathbf{p}_s = \bar{\phi}_s(\mathbf{x}_s) - \bar{\phi}_s(0)$  is the difference at  $t=s$ , then  $\Delta \mathbf{p}_t = (1-\gamma)\Delta \mathbf{p}_s$  for  $t>s$ , so that the permanent change is less than the immediate one. This establishes P2.

Consider next M6. Using the notation of the proposition and (37),

$$\bar{\phi}_{t+1}^*(0) = \delta \gamma \bar{\phi}_t^*(0) + \delta (1-\gamma) \mathbf{p}_t^*(\mathbf{x}_t^*) + (1-\delta) \bar{\phi}^0 \quad . \quad (41)$$

Subtracting (37) from this gives

$$\Delta \bar{\phi}_{t+1} = \delta \gamma \Delta \bar{\phi}_t + \delta (1-\gamma) \Delta \mathbf{p}_t \quad . \quad (42)$$

Up to the starting point  $t=s$ , marketing is the same in both scenarios and so  $\Delta \bar{\phi}_s = 0$ . At  $s$ ,  $\mathbf{x}_s^*$  induces some increment  $\Delta \mathbf{p}_s$  so that  $\Delta \mathbf{p}_{s+1} = \delta(1-\gamma)\Delta \mathbf{p}_s$ . Thereafter marketing is at its reference value,  $\mathbf{x}_t^* = \mathbf{x}_t = 0$ . Therefore, for  $t>s$ ,

$$\Delta \mathbf{p}_t = \mathbf{p}_t^*(0) - \mathbf{p}_t(0) = \bar{\phi}_t^*(0) - \bar{\phi}_t(0) = \Delta \bar{\phi}_t \quad . \quad (43)$$

so that  $\Delta \bar{\phi}_{t+1} = \delta \Delta \mathbf{p}_t$  and by recursion

$$\Delta \mathbf{p}_t = (1-\gamma) \delta^{t-s} \Delta \mathbf{p}_s \quad t>s \quad . \quad (44)$$

Over all  $t \geq s$ , the sum is

$$\frac{(1-\gamma \delta)}{1-\delta} \Delta \mathbf{p}_s \quad . \quad (45)$$

This completes P3.

## REFERENCES

- Allenby, Greg M. and Peter J. Lenk, (1992), "Modeling Household Purchase Behavior," WPS 92-39, College of Business, The Ohio State University, Columbus, Ohio 43210.
- Bucklin, Randolph E. and James M. Lattin, (1992), "A Model of Product Category Competition Among Grocery Retailers," *Journal of Retailing*, vol. 68, no. 3, 271-293.
- Chatfield, C., A.S.C. Ehrenberg and G.J. Goodhardt (1966), "Progress on a simplified model of stationary purchasing behavior," *J. of the Royal Statistical Society, A*, vol. 129, 316-337.
- Chintagunta, Pradeep K., Dipak C. Jain, and Naufel J. Vilcassim, (1991) "Investigating Heterogeneity in Brand Preferences in Logit Models for Panel Data," *Journal of Marketing Research*, November, 417-428.
- Davies, R.B. (1984), "A generalized beta-logistic model for longitudinal data with an application to residential mobility," *Environment and Planning A*, vol. 16, 1375-1386
- Dunn, Richard and Neil Wrigley (1985), "Beta-logistic models of urban shopping center choice," *Geographical Analysis*, vol. 17, no. 2, 95-113 (April).
- Fader, Peter S. (1993), "Integrating the Dirichlet-multinomial and multinomial logit models of brand choice," *Marketing Letters*, vol. 4, no. 2, 99-112 (April).
- Fader, Peter S. and James M. Lattin (1993), "Accounting for heterogeneity and non-stationarity in a cross-sectional model of consumer purchase behavior," *Marketing Science*, vol. 12, no. 3, 304-317 (Summer).
- Fader, Peter S., James M. Lattin and John D.C. Little (1992), "Estimating nonlinear parameters in the multinomial logit model," *Marketing Science*, vol. 11, no. 4, 372-385 (Fall).
- Gönül, Fusun, and Kannan Srinivasan (1993), "Modeling Multiple Sources of Heterogeneity in Multinomial Logit Models: Methodological and Managerial Issues," *Marketing Science*, vol. 12, no. 3, 213-229 (Summer).
- Goodhardt, G.J., A.S.C. Ehrenberg and C. Chatfield (1984), "The Dirichlet: a comprehensive model of buying behavior," *J. of the Royal Statistical Society, A*, vol. 147, Part 3, 621-655.
- Guadagni, Peter M. and John D.C. Little (1983), "A logit model of brand choice calibrated on scanner data," *Marketing Science*, vol. 2, no. 3, 203-238 (Summer).
- Heckman, James J. and Robert J. Willis (1977), "A beta-logistic model for the analysis of sequential labor force participation by married women," *J. of Political Economy*, vol. 85, no. 1, 27-58.

Jeuland, Abel P., Frank M. Bass and Gordon P. Wright (1980), "A multibrand stochastic model compounding heterogeneous Erlang timing with multinomial choice processes," *Operations Research*, vol. 28, 255-277 (March-April).

Krishnamurti, Lakshman and S.P. Raj (1988), "A model of brand choice and purchase quantity price sensitivities," *Marketing Science*, vol. 7, no. 1, 1-20 (Winter).

Lattin, James M. (1991), Measuring preference from scanner panel data: Filtering out the effects of price and promotion, Unpublished working paper, Graduate School of Business, Stanford University (January).

Little, John D.C. (1975), "BRANDAID: A marketing-mix model, Part 1: Structure," *Operations Research*, vol. 23, no. 2, 628-655 (July-August).

Pedrick, James H. and Fred S. Zufryden (1991), "Evaluating the impact of advertising media plans: a model of consumer purchase dynamics using single-source data," *Marketing Science*, vol. 10, no. 2, 111-130 (Spring).

Rossi, Peter E. and Greg M. Allenby (1993), "A Bayesian Approach to Estimating Household Parameters," *Journal of Marketing Research*, May, 171-182.

Samuels, Debra (1983), "The effectiveness of price-off label packages as a promotional tool for consumer packaged goods," S.M. Thesis, Sloan School of Management, M.I.T., Cambridge, MA 02139.

Uncles, Mark D. (1987), "A beta-logistic model of mode choice: goodness of fit and intertemporal dependence," *Transportation Research B*, vol. 21B, no. 3, 195-205.

Vilcassim, Naufel J. and Dipak C. Jain (1991), "Modeling Purchase Timing and Brand-Switching Behavior Incorporating Explanatory Variables and Unobserved Heterogeneity," *Journal of Marketing Research*, February, 29-41.