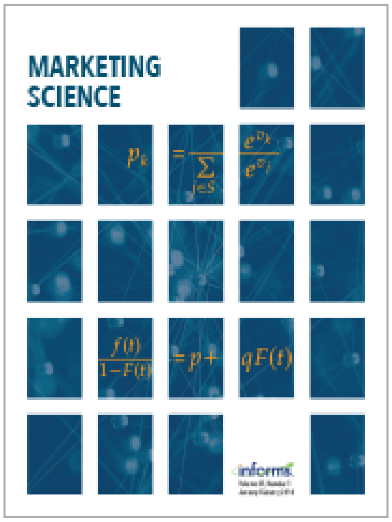



Marketing Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

# Customizing Promotions in Online Stores

Jie Zhang, Lakshman Krishnamurthi,

MARKETING SCIENCE
Vol. 23, No. 4, Fall 2004, pp. 561–578
ISSN 0732-2399 | EISSN 1526-548X | 04 | 2304 | 0561

DOI 10.1287/mksc.1040.0055


# Customizing Promotions in Online Stores

Jie Zhang
Stephen M. Ross School of Business, University of Michigan, 701 Tappan Street, Ann Arbor, Michigan 48109-1234, jiejie@umich.edu

Lakshman Krishnamurthi
Kellogg Graduate School of Management, Northwestern University, 2001 Sheridan Road, Evanston, Illinois 60208, laksh@kellogg.northwestern.edu

The main objective of this paper is to provide a decision-support system of micro-level customized promotions, primarily for use in online stores. Our proposed approach utilizes the one-on-one and interactive nature of the Internet shopping environment and provides recommendations on *when to promote how much to whom*. We address the issue by first constructing a joint purchase incidence-brand choice-purchase quantity model that incorporates how variety-seeking/inertia tendency differs among households and change over time for the same household. Based on the model, we develop an optimization procedure to derive the optimal amount of price discount for each household on each shopping trip. We demonstrate that the proposed customization method could greatly improve the effectiveness of current promotion practices, and discuss the implications for retailers and consumer packaged goods companies in the age of Internet technology.

*Key words*: customized promotions; profit optimization; Internet marketing; decision support system; personalized marketing; econometric models; purchase incidence; brand choice; purchase quantity; variety-seeking; inertia
*History*: This paper was received June 7, 2000, and was with the authors 21 months for 4 revisions, processed by Scott Neslin.

## 1. Introduction

Customized promotions have been gaining popularity in the retail industry as more companies come to realize the potential of customer-centric strategies over mass-market strategies (*Global Cosmetic Industry* December 2001). So far, industry practices and marketing academic research have mainly focused on the issue of how to *target* certain consumers for a given promotion, but the *depth* and *timing* of promotions are not tailored toward individuals nor adjusted by updated information on their purchases. On the other hand, technological development, especially the rapid growth of the Internet, has provided the potential to deliver promotions that are customized for each individual household on each shopping trip. This would represent micro-marketing at the finest level. The main objective of this research is to develop a decision-support system that provides recommendations on *when to promote how much to whom*, primarily for use in online stores.

Previous research on promotion decision-support systems has derived promotion calendars for brick-and-mortar retailers (Tellis and Zufryden 1995) or manufacturers (Silva-Risso et al. 1999) which are not differentiated at the individual household level. On the other hand, research addressing individual level customization topics has focused on targeting but not timing (e.g., Shaffer and Zhang 1995, Rossi et al. 1996). The emphasis of our research is on the *timing* of promotions. We believe that timing could make targeting more efficient and help achieve the full potential of micro-marketing, and that the Internet has provided an opportunity to realize such potential.

A distinctive feature of our proposed promotion customization system is that it utilizes the interactive and one-on-one nature of the Internet shopping environment. It derives the optimal price promotion for each household on each shopping trip by taking into account the time-varying pattern of purchase behavior and the impact of current promotion on future purchases. The promotion decision is updated on each subsequent shopping trip for a household.

We approach the problem by first constructing a joint purchase incidence-brand choice-purchase quantity model that incorporates how variety-seeking/inertia tendency may differ among households and change over time *within* the same household. Based on the proposed consumer response model, we develop a decision-support system to optimize the depth of promotion *for each household on each shopping trip*. Time-varying patterns of variety-seeking/inertia have a direct implication on how promotions should be timed. Most previous studies on variety-seeking and/or inertia assumed that these properties remain constant over time for the same individual/household (for some exceptions see

Kahn et al. 1986, Bawa 1990, Trivedi et al. 1994). Timing would play a more important role in promotion decisions if the variety-seeking/inertia tendency varies over time for a substantial proportion of consumers in a market. For example, a promotion aimed at a consumer who did not buy the target brand on the previous occasion may not be profitable if she is in a high inertia state, while a promotion aimed at a consumer who bought the target brand on the previous occasion may not be profitable if she is in a high variety-seeking state. Our empirical analysis shows that the variety-seeking/inertia tendency does vary over time for the majority of households in the data. Therefore, we believe that a decision-support system designed to offer customized promotions should take into account such behavioral changes.

The rest of this paper is organized as follows. In §2, we describe the formulation of our consumer response model. In §3, we develop an optimization procedure for customized promotions. In §4, we present model estimation results and demonstrate the application and effectiveness of the proposed customization method. In §5, we discuss the limitations of this research and managerial implications for retailers and consumer packaged goods manufacturers.

## 2. Model Formulation

Previous research has demonstrated that it is important to take into account the interdependence in purchase incidence, brand choice, and purchase quantity decisions (e.g., Chiang 1991, Chintagunta 1993, Arora et al. 1998, Bell et al. 1999). We develop a new formulation to model the three components *simultaneously*. Our model shares similarities with those in the studies cited above, yet has unique features in certain aspects. Define

$I_{it} = 1$ if household $i$ makes a category purchase at shopping trip $t$; 0 otherwise,

$B_{ikt} = 1$ if household $i$ purchases alternative $k$ at shopping trip $t$; 0 otherwise,

$Q_{ikt} =$ household $i$'s purchase quantity of alternative $k$ at shopping trip $t$.

We derive the joint probability $\Pr(I_{it} = 1, B_{ikt} = 1, Q_{ikt} = q)$ as follows.

### 2.1. Purchase Incidence and Brand Choice

The utility function of alternative $k$ at shopping trip $t$ for household $i$ is given by

$$U_{ikt} = V_{ikt} + \varepsilon_{ikt} = X_{kt}\beta_i + \varepsilon_{ikt}, \quad k = 1, \ldots, K, \qquad (1)$$

where $V_{ikt}$, the systematic component of brand utilities, is a function of brand-specific constants, marketing-mix variables such as regular price and price discount, and a time-varying purchase event feedback effect component. The specification of $V_{ikt}$ will be elaborated later.

At each shopping trip $t$, the shopper decides whether to make a category purchase and will do so only if the utilities of the alternatives under consideration exceed a threshold. We formulate the threshold as

$$U_{i0t} = V_{i0t} + \varepsilon_{i0t} = \lambda_{0i} + \lambda_{1i} FREQ_i + \lambda_{2i} LQ_{it} + \varepsilon_{i0t}, \quad (2)$$

where $\lambda_{0i}$ is a constant, $FREQ_i$ is household $i$'s purchase frequency in the initialization period, and $LQ_{it}$ is the household's mean-centered last purchase quantity before shopping trip $t$. $LQ_{it}$ is used to capture inventory effects on purchase incidence decisions. To control for difference in purchase quantity across households, $LQ_{it}$ is computed as the last purchase quantity minus the household's average purchase quantity in the initialization period. This operationalization has been used by Jain and Vilcassim (1991) and Chintagunta and Haldar (1998). Note that it would not be appropriate to include an inventory variable in our model because it requires the use of the interpurchase duration which is endogenous to purchase incidence decisions (Chintagunta and Haldar 1998).

Assume that $\varepsilon_{ikt}$, $k = 0, 1, \ldots, K$, follows i.i.d. Type I extreme value distribution with location parameter 0 and scale parameter 1. It can be shown that the category purchase incidence probability and conditional brand choice probability are

$$\Pr(I_{it} = 1) = \Pr\Big(\max_j \{V_{ijt} + \varepsilon_{ijt}\} > V_{i0t} + \varepsilon_{i0t}\Big) = \frac{\sum_{j=1}^{K} \exp(V_{ijt})}{\exp(V_{i0t}) + \sum_{j=1}^{K} \exp(V_{ijt})} \quad \text{and} \qquad (3)$$

$$\Pr(B_{ikt} = 1 \mid I_{it} = 1) = \frac{\exp(V_{ikt})}{\sum_{j=1}^{K} \exp(V_{ijt})}, \qquad (4)$$

respectively.[1] Note that because $V_{ikt}$ is a function of price and promotion, among other variables, a promotion for any brand will increase the category purchase incidence probability. The joint probability of purchase incidence and choice is

$$\Pr(I_{it} = 1, B_{ikt} = 1) = \frac{\exp(V_{ikt})}{\exp(V_{i0t}) + \sum_{j=1}^{K} \exp(V_{ijt})}. \qquad (5)$$

[1] Our formulation can be seen as a special case of the nested logit model developed by McFadden (1978), in which the inclusive value is fixed to 1. As will be explained later, our model does not suffer from the Independence of Irrelevant Alternatives (IIA) problem because it includes a distance measure to accommodate similarities among choice alternatives. We estimated the nested logit inclusive value parameter in a latent-class incidence + choice model and found it to be almost 1 in our model formulation.

### 2.2. Purchase Quantity

Let $Q^*_{ikt}$ be a latent variable that determines how much household $i$ wants to buy alternative $k$ at shopping trip $t$, and $Q_{ikt}$ be the observed purchase quantity. Then,

$$Q_{ikt} = \begin{cases} Q^*_{ikt} & \text{if } I_{it} = 1 \text{ and } B_{ikt} = 1, \\ 0, & \text{otherwise.} \end{cases}$$

Specify

$$Q^*_{ikt} = Z_{ikt}\phi_i + \xi_{ikt} = \phi_{0i} + \phi_{1i} AQ_i + \phi_{2i} FREQ_i + \phi_{3i} RP_{kt} + \phi_{4i} PC_{kt} + \xi_{ikt}, \quad (6)$$

where $AQ_i$ is household $i$'s average purchase quantity and $FREQ_i$ is its purchase frequency in the initialization period, $RP_{kt}$ and $PC_{kt}$ are alternative $k$'s regular price and price cut, and $\xi_{ikt}$ is the unobserved random term.[2]

The interdependence between purchase incidence and choice and quantity decisions is formulated as follows:

$$\Pr(I_{it}=1, B_{ikt}=1) = \Pr\left(V_{ikt} + \varepsilon_{ikt} > \max_{\substack{j=0,1,\dots K \\ \text{and } j\neq k}} \{V_{ijt} + \varepsilon_{ijt}\}\right) = \Pr\left(\max_{\substack{j=0,1,\dots K \\ \text{and } j\neq k}} \{V_{ijt} + \varepsilon_{ijt}\} - \varepsilon_{ikt} < V_{ikt}\right).$$

Let $\varepsilon^*_{ikt} = \max_{j=0,1,\dots K \text{ and } j\neq k}\{V_{ijt} + \varepsilon_{ijt}\} - \varepsilon_{ikt}$. By properties of the extreme value distribution (see Ben-Akiva and Lerman, 1985, p. 105), $\max_{j=0,1,\dots K \text{ and } j\neq k}\{V_{ijt} + \varepsilon_{ijt}\}$ follows a Type I extreme value distribution with location parameter $\ln[\sum_{j=0,1,\dots K \text{ and } j\neq k} \exp(V_{ijt})]$ and scale parameter 1, and $\varepsilon^*_{ikt}$ follows a logistic distribution with cumulative distribution function (CDF):

$$F(\varepsilon^*_{ikt}) = \frac{1}{1 + \exp\{\ln[\sum_{j=0,1,\dots K \text{ and } j\neq k} \exp(V_{ijt})] - \varepsilon^*_{ikt}\}}.$$

Assume that the quantity random term, $\xi_{ikt}$, follows a logistic distribution with mean 0 and scale parameter $\delta_\xi$. Its CDF is given by

$$F(\xi_{ikt}) = \frac{1}{1 + \exp(-\delta_\xi \xi_{ikt})},$$

and its variance is $\sigma^2_\xi = \pi^2/3\delta_\xi$. We adopt a flexible bivariate logistic distribution proposed by Gumbel (1961) for the joint distribution of $\varepsilon^*_{ikt}$ and $\xi_{ikt}$. Unlike the standard bivariate logistic distribution in which the correlation between the two variables is fixed to 0.5, this formulation allows the correlation coefficient to be estimated from the data. The joint CDF is given by (Gumbel 1961, p. 347)

$$F(\varepsilon^*_{ikt}, \xi_{ikt}) = F(\varepsilon^*_{ikt})F(\xi_{ikt}) \cdot [1 + \theta(1 - F(\varepsilon^*_{ikt}))(1 - F(\xi_{ikt}))], \quad -1 \le \theta \le 1, \quad (7)$$

where $\theta$ is to be estimated from the data. The correlation coefficient of $\varepsilon^*_{ikt}$ and $\xi_{ikt}$ is $\rho = 3\theta/\pi^2$. To ensure that the estimate of $\theta$ falls between $[-1, 1]$, we use $\theta = \sin(\tau)$ in the estimation procedure.

The probability of observing $Q_{ikt} > 0$ is

$$\Pr(I_{it} = 1, B_{ikt} = 1, Q_{ikt} = q_{ikt}) = \Pr(\varepsilon^*_{ikt} < V_{ikt}, Z_{ikt}\phi_i + \xi_{ikt} = q_{ikt}) = \int_{-\infty}^{V_{ikt}} f(\varepsilon^*_{ikt}, \xi_{ikt} = q_{ikt} - Z_{ikt}\phi_i)\, d\varepsilon^*_{ikt},$$

where $f(\cdot,\cdot)$ is the joint density function of $\varepsilon^*_{ikt}$ and $\xi_{ikt}$. The bivariate logistic distribution enables us to get a closed-form expression of the above integral. We present the result here and show the derivation in Appendix C:

$$\Pr(I_{it} = 1, B_{ikt} = 1, Q_{ikt} = q_{ikt}) = \frac{e^{V_{ikt}}}{\sum_{j=0}^{K} e^{V_{ijt}}} \frac{\delta_\xi e^{\delta_\xi (Z_{ikt}\phi_i - q_{ikt})}}{[1 + e^{\delta_\xi (Z_{ikt}\phi_i - q_{ikt})}]^2} \cdot \left[1 + \theta\left(1 - \frac{e^{V_{ikt}}}{\sum_{j=0}^{K} e^{V_{ijt}}}\right) \frac{-1 + e^{\delta_\xi (Z_{ikt}\phi_i - q_{ikt})}}{1 + e^{\delta_\xi (Z_{ikt}\phi_i - q_{ikt})}}\right] \quad (8)$$

In Equation (8), $e^{V_{ikt}}/\sum_{j=0}^{K} e^{V_{ijt}}$ is the joint probability $\Pr(I_{it} = 1, B_{ikt} = 1)$, $\delta_\xi e^{\delta_\xi (Z_{ikt}\phi_i - q_{ikt})}/[1 + e^{\delta_\xi (Z_{ikt}\phi_i - q_{ikt})}]^2$ is the marginal probability $\Pr(Q_{ikt} = q_{ikt})$, and $\theta$ captures the interdependence of $\varepsilon^*_{ikt}$ and $\xi_{ikt}$. When $\theta = 0$, Equation (8) reduces to $\Pr(I_{it} = 1, B_{ikt} = 1, Q_{ikt} = q_{ikt}) = \Pr(I_{it} = 1, B_{ikt} = 1)\Pr(Q_{ikt} = q_{ikt})$. In other words, $\Pr(Q_{ikt} = q_{ikt} \mid I_{it} = 1, B_{ikt} = 1) = \Pr(Q_{ikt} = q_{ikt})$. Because the choice probability $\Pr(I_{it} = 1, B_{ikt} = 1)$ decreases with $\varepsilon^*_{ikt}$ by the definition of $\varepsilon^*_{ikt}$, $\theta > 0$ indicates that unobserved factors in the quantity and choice components for alternative $k$ are *negatively* correlated, and $\theta < 0$ indicates that unobserved factors in the two components are *positively* correlated. In the extreme case where $\Pr(I_{it} = 1, B_{ikt} = 1) = 1$, i.e., if alternative $k$ is always purchased, Equation (8) reduces to $\Pr(I_{it} = 1, B_{ikt} = 1, Q_{ikt} = q_{ikt}) = \Pr(Q_{ikt} = q_{ikt})$, regardless of whether $\varepsilon^*_{ikt}$ and $\xi_{ikt}$ are correlated.

[2] We also tested $LQ_{it}$ in the latent quantity equation. It turned out to be insignificant in our empirical estimation. In addition, $FREQ_i$ was insignificant in the latent quantity equation and was dropped in the final model for the butter category.

### 2.3. Brand Utility Functions

We now explain the brand utility functions in detail with an emphasis on the *purchase event feedback effect*, which is defined as the impact of past purchases on current brand preference (cf., Gedenk and Neslin 1999). We distinguish between two types of purchase event feedback effects. The first type is attributed to a consumer's intentional tendency to stick to or switch away from an item bought on the previous category purchase occasion, which is not caused by external influences such as a price discount. This phenomenon is here referred to as *inertia* or *variety-seeking*, respectively, and has been studied extensively in the marketing literature (e.g., Jeuland 1979, Lattin 1987, Bawa 1990, Seetharaman and Chintagunta 1998). The second type is the promotion-related purchase event feedback effect.[3] The impact of past purchase on a consumer's current preference for a brand may depend on whether it was purchased on promotion. Gedenk and Neslin (1999) have found that *price* promotions are associated with a negative effect on the impact of past purchase on current brand preference. We incorporate both types of purchase event feedback effects and represent the brand utility function in the following general expression:

$$U_{ikt} = \alpha_{ki} + X_{kt}\beta_i + \gamma_{it} LB_{ikt} + \omega_i LB_{ikt} LPROM_{ikt} - \gamma_{it} d_{kI_i} + \varepsilon_{ikt}, \tag{9}$$

where $\alpha_{ki}$ is the alternative specific constant for household $i$, $X_{kt}$ is a vector of marketing-mix variables, $LB_{ikt}$ equals 1 if alternative $k$ was chosen by household $i$ on the last purchase occasion prior to shopping trip $t$ and 0 otherwise, $LPROM_{ikt}$ equals 1 if alternative $k$ was on price promotion on household $i$'s last purchase occasion prior to shopping trip $t$ and 0 otherwise, and $d_{kI_i}$ is a measure of distance between alternatives $k$ and $I_i$, the alternative chosen by household $i$ at the previous purchase occasion, and $d_{kk} = 0$ by definition. For the alternative that was bought on the previous occasion, say $k$, Equation (9) simplifies to

$$U_{ikt} = \alpha_{ki} + X_{kt}\beta_i + \gamma_{it} + \omega_i LPROM_{ikt} + \varepsilon_{ikt}, \tag{9a}$$

and for any other alternative $j \neq k$, Equation (9) simplifies to

$$U_{ijt} = \alpha_{ji} + X_{jt}\beta_i - \gamma_{it} d_{kj} + \varepsilon_{ijt}. \tag{9b}$$

As shown in Equation (9a), parameter $\gamma_{it}$ indicates the change in a brand's utility because of being purchased previously, and thus captures a household's variety-seeking/inertia tendency, with $\gamma_{it} > 0$ indicating inertia, $\gamma_{it} < 0$ indicating variety-seeking, and $\gamma_{it} = 0$ indicating zero-order behavior. We allow $\gamma_{it}$ to vary over time for a given household and will provide the details shortly. Parameter $\omega_i$ captures the impact of price promotion on purchase event feedback. $\omega_i$ should be negative if a *price* promotion induced purchase is associated with a lower purchase event feedback effect compared to a purchase not on promotion, as documented by Gedenk and Neslin (1999). The term $(-\gamma_{it} d_{kj})$ in Equation (9b) is included to modify the impact of variety-seeking/inertia on the other alternatives $j \neq k$ due to the differences in similarity among choice alternatives. It indicates that the utility of an alternative at a given time depends on the consumer's variety-seeking/inertia tendency as well as its (dis)similarity to the previously purchased item. Like some other models of variety-seeking/inertia (e.g., Papatla and Krishnamurthi 1992), our model is capable of capturing the phenomenon that an inertial consumer is more likely to repeat the brand she purchased last time and more likely to switch to something similar if she does switch, and that a variety-seeking consumer is more likely to switch to another alternative, with the probability being higher for the most dissimilar item. Distance $d_{kj}$ can either be estimated as parameters in the model or assume predetermined values based on external sources of information. In this study, we adopt the latter approach and conducted a survey to compute the distance measures (see Appendix A for details).[4]

### 2.4. Time-Varying Patterns of Variety-Seeking and Inertia

Consistent with previous findings that variety-seeking or inertia tendency may not be constant for the same individual over time (Bawa 1990, Seetharaman et al. 1999), we allow the parameter $\gamma_{it}$ to change over $t$ for each $i$. There are advantages and disadvantages of variety-seeking/inertia to a consumer. Variety-seeking breaks the routine and brings novelty and excitement to the consumption experience. Yet, it also demands more time and cognitive effort and could be tiring if a consumer repeatedly seeks variety. On the other hand, inertia saves her from spending too much cognitive effort, but it could build satiation and cause

[3] We thank the area editor for suggesting the promotion-related effect.

[4] We also tested the first option by estimating the distance measures along with other parameters from the data. This approach provided a slightly better fit to the data, while performing worse in the holdout sample prediction in our empirical analysis. The general pattern of the other parameter estimates remains the same using the two approaches except that a few parameters are less stable in the model that estimates the distance parameters. (Details are available from the authors upon request.) Because our main research objective is to utilize the consumer response model to construct an optimization system for which holdout sample predictability is more important, we opted to use the survey-based distance measure. We thank an anonymous reviewer for suggesting this approach.

boredom after a while. Therefore, inertia may first increase with repeat purchases when the benefit of inertia dominates. As satiation builds up her inertia tendency may start declining over time until the negative side of inertia dominates and she switches into a variety-seeking state. Using the same logic, her variety-seeking tendency may increase with the number of switches initially and then decrease after a certain point until she switches back to inertia. This leads to the following formulation:[5]

$$\gamma_{it_m} = \gamma_{iT_{m-1}} + \mu_{1i} L_{1i,\,t_m-1} + \eta_{1i} L^2_{1i,\,t_m-1} \quad \text{when} \quad L_{1i,\,t_m-1} > 0 \quad \text{and} \quad L_{2i,\,t_m-1} = 0, \tag{10a}$$

$$\gamma_{it_m} = \gamma_{iT_{m-1}} + \mu_{2i} L_{2i,\,t_m-1} + \eta_{2i} L^2_{2i,\,t_m-1} \quad \text{when} \quad L_{2i,\,t_m-1} > 0 \quad \text{and} \quad L_{1i,\,t_m-1} = 0, \tag{10b}$$

where $m$ indicates the $m$th sequence of consecutive purchases of either the same alternative in a row or different alternatives in a row, $L_{1i,\,t_m-1}$ indicates the number of times household $i$ has been buying the same alternative *consecutively in the absence of promotion* up to shopping trip $t_m - 1$, and $L_{2i,\,t_m-1}$ indicates the number of times household $i$ has been buying different alternatives *consecutively in the absence of promotion* up to shopping trip $t_m - 1$. By definition, one of these variables is greater than one and the other is zero on all shopping trips in the estimation data. Note that $\gamma_{it_m}$ is updated in a continuous manner from the previous run by having the term $\gamma_{iT_{m-1}}$ on the right-hand side of the equations.

The variables $L_{1i,\,t}$ and $L_{2i,\,t}$[6] are constructed as proxies for the number of purchase occasions that a consumer has been engaged in inertia or variety-seeking, respectively. Consistent with the literature, a key feature of our operational definitions of variety-seeking/inertia is the distinction between switches/repeats driven by variety-seeking/inertia and those induced by external factors. We focus on one external factor, promotion, because of its managerial importance and availability in most purchase data, although the logic can be applied to other external factors. We attribute a switch/repeat *in the absence of promotions* to variety-seeking/inertia. If the chosen item is on promotion, we assume that the purchase is at least partly driven by the promotion and it does not provide evidence of a consumer being in an variety-seeking/inertia state.[7]

Both $L_{1i,\,t-1}$ and $L_{2i,\,t-1}$ are set to zero at the beginning of the data for each household.[8] The household's initial variety-seeking/inertia tendency is captured by the baseline parameter $\gamma_{i0}$. Every time a household purchases an item on promotion, $L_{1i,\,t}$ and $L_{2i,\,t}$ remain at their values on the previous purchase occasion. Every time a household repeats the same item in the absence of a promotion, $L_{1i,\,t}$ increases by one and $L_{2i,\,t}$ is set to zero. Every time a household makes a switch in the absence of a promotion, $L_{2i,\,t}$ increases by one and $L_{1i,\,t}$ is set to zero. The following example illustrates the computation of $L_{1i,\,t}$ and $L_{2i,\,t}$. Quadratic forms of $L_{1i,\,t-1}$ and $L_{2i,\,t-1}$ are employed to capture possible nonlinear patterns of state dependence.[9]

| Purchase occasion | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Item chosen | A | A | A | A | B | C | B | D | D | D |
| On promotion? | No | No | Yes | No | No | No | Yes | No | No | No |
| $L_1$ | 0 | 1 | 1 | 2 | 0 | 0 | 0 | 0 | 1 | 2 |
| $L_2$ | 0 | 0 | 0 | 0 | 1 | 2 | 2 | 3 | 0 | 0 |

To summarize, the impact of price promotion on purchase event feedback is captured by parameter $\omega_i$, while the variety-seeking/inertia parameter $\gamma_{it}$ captures the nonpromotion related purchase event feedback effect. We would expect the following signs of the parameters if the negative effect of price promotion on purchase event feedback and the cyclical pattern of time-varying inertia and variety-seeking, as previously discussed, hold in the data: $\omega_i$ negative, $\mu_{1i}$ positive, $\eta_{1i}$ negative, $\mu_{2i}$ negative, and $\eta_{2i}$ positive.

To our knowledge, Bawa's (1990) hybrid model was the first nonstochastic model to capture time-varying patterns of inertia and variety-seeking. The model allows a consumer to exhibit a hybrid behavior of inertia and variety-seeking at different times during her purchase history. The hybrid model assumes that a consumer's state dependence gets renewed every time a brand switch occurs. This renewal process assumption implies that choices made prior to the most recent brand switch do not influence current purchase behavior. It is unlikely, however, that actual purchase behavior experiences such a sudden

[5] Seetharaman et al. (1999) find that variety-seeking/inertia also diminish with interpurchase duration based on their brand choice model. Because interpurchase duration is endogenous to purchase incidence decisions, we do not include this variable in our model. In addition, we tested our model with interpurchase duration parameters (one for variety-seeking and another for inertia) versus without them and found that the difference in the log-likelihood was not significant. We thank an anonymous reviewer for pointing out the endogeneity issue.

[6] The subscript $m$ in $t_m$ is omitted for notation simplicity.

[7] $L_{1,\,t-1}$ and $L_{2,\,t-1}$ are used to predict $\gamma_t$. Like any other model, the prediction is not perfect. It is possible to have ($L_{1t} > 0$ and $\gamma_t < 0$) and ($L_{2t} > 0$ and $\gamma_t > 0$) in an empirical application. Nevertheless, this situation occurred rarely in our application (1.3% of observations for butter and 0.7% of observations for liquid detergent). We thank the area editor for pointing out this issue.

[8] The estimation data do not include the first observation for each household.

[9] Log-transformations of $L_1$ and $L_2$ are tested and found to be inferior to the quadratic form in our application.

discontinuity after a switch. Moreover, in the case of frequent brand switches, the assumption implies that purchases made one or two occasions earlier do not have any impact on the current decision. Our model does not impose this renewal process assumption.

Another feature of our variety-seeking/inertia formulation is the distinction between switches/repeats made on promotion and not on promotion. We allow these occasions to be driven by different forces. In contrast, the hybrid model treats purchases made on and off promotion in the same manner. Further, in the hybrid model the utilities of the alternatives not chosen at the previous occasion are assumed to be the same, while they are allowed to be different depending on the variety-seeking/inertia parameter and the distance measure in our model. In addition, our model incorporates the effect of marketing-mix variables while the hybrid model does not.

### 2.5. Consumer Heterogeneity and the Log-Likelihood Function

The model has been constructed at the individual household level so far. We adopt a latent class model to capture unobserved consumer heterogeneity (cf., Kamakura and Russell 1989), in which parameters are segment-specific, denoted by the subscript $g = 1, \ldots, G$. The log-likelihood function is given by

$$LL = \sum_{i=1}^{N} \log\left( \sum_{g=1}^{G} q_g \prod_{t=1}^{T_i} \Pr_g(I_{it} = 0)^{1-I_{it}} \cdot \prod_{k=1}^{K} \Pr_g(I_{it} = 1, B_{ikt} = 1, Q_{ikt} = q_{ikt})^{I_{it} * B_{ikt}} \right), \quad (11)$$

where $q_g$ is the probability of belonging to segment $g$, and other terms are as shown previously. The number of latent segment $G$ is determined empirically by comparing the Bayesian Information Criterion (BIC) of models with different $G$ and the one that yields the lowest BIC is selected.

In terms of the way to model purchase incidence/brand choice and purchase quantity decisions jointly, our approach is similar to the ones in Krishnamurthi and Raj (1988)[10] and Bell et al. (1999), who assumed pairwise bivariate *normal* distributions of $\varepsilon^*_{ikt}$ and $\xi_{ikt}$ or a transformation of them. Because there is no closed-form expression of the integral based on the bivariate normal distribution, Bell et al. (1999) used numerical integration in the estimation, while Krishnamurthi and Raj (1988) used a two-stage estimation method which yielded consistent, but not efficient, parameter estimates. In addition, neither model takes into account unobserved heterogeneity in consumers' preference and responses to marketing mix. It would have been computationally formidable to combine unobserved heterogeneity and correlations of $\varepsilon^*_{ikt}$ and $\xi_{ikt}$ based on their formulations of bivariate-normal-distributed error terms. By utilizing the bivariate logistic distribution described above, we are able to not only get a closed-form likelihood function and thus simplify the estimation procedure, but also incorporate heterogeneity in the standard finite mixture model framework.

Hanemann (1984), Chiang (1991), Chintagunta (1993), and Arora et al. (1998) modeled the incidence/choice and quantity decisions jointly by deriving the components from a single utility maximization framework. They utilized properties of the extreme value distributions to obtain the conditional distribution of the quantity errors given a particular brand being chosen. Their approach takes into account the interdependence of the incidence, choice, and quantity decisions without explicitly specifying correlation parameters. These models assume that quantity is a direct function of the brand utility, which implies "a very direct type of substitution between quantity and quality in the consumer's preferences," a restriction pointed out by Hanemann (1984, p. 552). In our model, only the incidence/choice part is derived from utility maximization. An advantage of our approach (as well as those in Krishnamurthi and Raj 1988 and Bell et al. 1999) is that it does not impose this restriction and, thus, the model allows the variables and their coefficients in the quantity equations to be different from those in the brand utility and category threshold equations.

## 3. Customizing Promotions

The goal of our customization method is to derive the optimal promotion for each household on each shopping trip. To keep the derivation tractable, we choose to focus on price promotions, although the methodology can be well applied to other promotion variables.

We propose the following mechanism to implement our customized promotions. Assume an (online) retailer offers this special customization service to its clients. There should be only one client for this service within a product category. The retailer is free to accept other regular trade promotions. The customized promotions are sponsored by a manufacturer and implemented by the retailer. The retailer is compensated for administering the promotions and also gets a share of the incremental profits generated by the service. The retailer essentially works as the manufacturer's agent and they have a common profit objective. This sole manufacturer sponsor model has been used in practice. For example, Catalina Marketing Corporation's in-store coupon distribution service

[10] Krishnamurthi and Raj (1988) did not model purchase incidence.

Checkout Coupon® offers one manufacturer the right to be the sponsor of a product category for competitive coupons. A sole manufacturer sponsor model minimizes competition among manufacturers within a category. In addition, it increases the retailer's incentive to carefully monitor the implementation process because the extra profit is to be shared by the manufacturer and the retailer.

### 3.1. Objective Function

A manufacturer wants to optimize its promotions by taking into account their impact on future purchases. The time horizon of this dynamic optimization depends on how often the manufacturer is able and willing to update its promotion decisions. In an online context, the store interface can be modified for each individual household every time the household logs on to the store and promotion decisions can be updated using its purchase history data up until the previous visit. Therefore, the time horizon for the dynamic optimization can be much shorter than in the case of traditional promotion calendar planning in brick-and-mortar stores. We choose to use three time periods for exposition purposes. While this is a reasonable choice from a practical point of view, it can be easily modified to fit a manufacturer's specific needs. When manufacturers or retailers optimize promotion scheduling in brick-and-mortar stores, they usually need to consider the entire time period, often three months, in advance (e.g., Tellis and Zufryden 1995, Silva-Risso et al. 1999). A major advantage of our proposed promotion method over traditional ones is that decisions can be updated based on the most recent *actual* purchase behavior. This avoids the need to predict what would happen during the entire time period, as well as having to make decision recommendations based on such long-term predictions.

Assume that a manufacturer wants to optimize its expected gross profit from a household over three shopping trips with respect to a brand's price promotions at the three trips, $PC_{kt}$, $PC_{k,t+1}$, and $PC_{k,t+2}$. The expected profit over three shopping trips is computed using parameter estimates of our model. The objective function is

$$\max_{PC_{k,t+s},\, s=0,1,2} E(\pi_{ikt}) = \max_{PC_{k,t+s},\, s=0,1,2} \left\{ \sum_{s=0}^{2} \hat{P}(I_{i,t+s}=1, B_{ik,t+s}=1) \cdot E(Q_{ik,t+s} \mid I_{i,t+s}=1, B_{ik,t+s}=1) \cdot (M_{k,t+s} - PC_{k,t+s}) \right\}, \tag{12}$$

where $k$ is the target brand, $E(Q_{ik,t+s} \mid I_{i,t+s}=1,$ $B_{ik,t+s}=1)$, $s=0,1,2$, are expected purchase quantities of $k$ given that $k$ is chosen at $t, t+1$, and $t+2$, $M_{kt}$ is the manufacturer's regular profit margin of $k$ without a promotion at $t$, $\hat{P}(I_{i,t+s}=1, B_{ik,t+s}=1)$ are predicted joint probabilities of purchase incidence and brand choice at $t$, $t+1$, and $t+2$, and are functions of the decision variables $PC_{kt}$, $PC_{k,t+1}$, and $PC_{k,t+2}$, as well as other marketing-mix variables in the model.

The optimization is subject to the following constraints:

(1) $PC_{k,t+s} \geq 0$, $s=0,1,2$;

(2) $PC_{k,t+s} \leq b_k RP_{k,t+s}$, $s=0,1,2$, where $b_k \in (0,1)$ is the fraction of brand $k$'s regular price for a maximum price discount. In practice manufacturers often impose an upper limit for the depth of price discounts for fear of brand equity erosion. $b_k$ is prespecified by the manufacturer.

In actual business practice, $M_{kt}$, the manufacturer's regular profit margin of brand k, is known to the manufacturer as well as the retailer via the cooperation in a customized promotion program. Its value, however, is unavailable in our data. In the application section, we assume values for the manufacturer and retailer margins, denoted by $m_m$ and $m_r$, respectively, to derive $M_{kt}$ from regular price $RP_{kt}$. It is easy to see that $M_{kt} = m_m(1-m_r)RP_{kt}$.

The expected quantity conditional on a brand being purchased is derived based on the probability density function of $P(Q_{ikt} \mid I_{it}=1, B_{ikt}=1)$. We present the result here and show the derivation in Appendix D:

$$E(Q_{ikt} \mid I_{it}=1, B_{ikt}=1) = \frac{1}{\delta_\xi}\left[\log(1+e^{\delta_\xi Z_{ikt}\phi_i}) - \theta\left(1 - e^{V_{ikt}} \Big/ \sum_{j=0}^{K} e^{V_{ijt}}\right) \cdot \frac{e^{\delta_\xi Z_{ikt}\phi_i}}{1+e^{\delta_\xi Z_{ikt}\phi_i}}\right]. \tag{13}$$

The joint probabilities $P(I_{i,t+s}=1, B_{ik,t+s}=1)$ depend on what brand was chosen on the previous purchase occasion (see Equations (9a) and (9b)). At each shopping trip $t$, the last brand purchased variable $LB_{kt}$ is known, while $LB_{k,t+1}$ and $LB_{k,t+2}$ are not. We obtain $\hat{P}(I_{it}=1, B_{ikt}=1)$ using $LB_{kt}$. A "tree-structure" of possible outcomes (i.e., purchase/no purchase, and brand chosen if a purchase is made) at $t$ and $t+1$ is utilized to compute $\hat{P}(I_{i,t+1}=1, B_{ik,t+1}=1)$ and $\hat{P}(I_{i,t+2}=1, B_{ik,t+2}=1)$ by using the *predicted* conditional probabilities, each conditional on the outcome on the previous shopping trip. The values of $LB_{k,t+1}$ and $LB_{k,t+2}$ are determined based on the tree structure. Specifically, the predicted joint probabilities are computed as follows:

$$\hat{P}(I_{i,t+1}, B_{ik,t+1}) = \hat{P}(I_{it}=0)\hat{P}(I_{i,t+1}, B_{ik,t+1} \mid I_{it}=0) + \sum_{j=1}^{K} \hat{P}(I_{it}=1, B_{ijt}=1) \cdot \hat{P}(I_{i,t+1}, B_{ik,t+1} \mid I_{it}=1, B_{ijt}=1), \quad (14)$$

$$\hat{P}(I_{i,t+2}, B_{ik,t+2}) = \hat{P}(I_{i,t+1}=0)\hat{P}(I_{i,t+2}, B_{ik,t+2} \mid I_{i,t+1}=0) + \sum_{j=1}^{K} \hat{P}(I_{i,t+1}=1, B_{ij,t+1}=1) \cdot \hat{P}(I_{i,t+2}, B_{ik,t+2} \mid I_{i,t+1}=1, B_{ij,t+1}=1). \quad (15)$$

Note that we also need to derive the variety-seeking/inertia parameters $\gamma_{t+1}$ and $\gamma_{t+2}$ when computing the above equations. $\gamma_{t+1}$ and $\gamma_{t+2}$ depend on the tree-structure of possible purchase outcomes and whether a brand is purchased on promotion. The combination determines the values of variables $L_1$ and $L_2$, which in turn determine the values of $\gamma_{t+1}$ and $\gamma_{t+2}$ according to Equations (10a) and (10b). (See Appendix B for the details.)

It is obvious that $\hat{P}(I_{i,t+1}=1, B_{ik,t+1}=1)$ is a function of $PC_{kt}$ and $PC_{k,t+1}$, and $\hat{P}(I_{i,t+2}=1, B_{ik,t+2}=1)$ is a function of $PC_{kt}$, $PC_{k,t+1}$, and $PC_{k,t+2}$. Thus, our formulation incorporates the notion that current promotions affect future purchases. For each household on a given shopping trip $t$, the optimal price promotions for $t, t+1$ and $t+2$ can be derived using purchase data up to $t-1$. On the household's next shopping trip $t+1$, the promotion decision is derived for $t+1$, $t+2$, and $t+3$ using purchase data up to $t$. Therefore, the proposed optimization can be performed for each household on every shopping trip using its most recent purchase data.

Note that, in our formulation, the impact of current promotions on future purchases are incorporated in the *predicted* conditional probabilities and only lagged purchase variables *prior to shopping trip t* enter the objective function. Some previous dynamic optimization models in the marketing literature (e.g., Tellis and Zufryden 1995) used the *actual* lagged choice variables at all time periods in the objective function. This is problematic for a couple of reasons: First, in reality one never has data on future outcomes; and second, the very idea of current promotion affecting future purchases is contradicted because it is assumed that the future choice outcomes would be what they were in the data as opposed to being affected by the promotion decision. Our formulation overcomes these problems.

Because a brand's utility function depends on whether it was chosen previously, we need to distinguish two types of shopping trips when implementing the optimization routine: those when a consumer purchased the target brand on the previous purchase occasion, and those when a consumer did not purchase the target brand on the previous purchase occasion. For ease of exposition, we refer to the first type of shopping trips as *previous-period buyers* and the second type of shopping trips as *previous-period nonbuyers*. For previous-period buyers, Equation (9a) applies to the target brand and Equation (9b) applies to all other brands. For previous-period nonbuyers, Equation (9a) applies to the brand chosen on the previous purchase occasion, and Equation (9b) applies to the target brand and all the other brands. In other words, the computation of $\hat{P}(I_{i,t}=1, B_{ik,t}=1)$ differs between the two cases. At $t+1$ and $t+2$, the computation will be the same for the two cases because we use predicted conditional probabilities conditional on all possible purchase outcomes on the previous trip.

We summarize the optimization procedure for deriving the customized price promotions during a given time period for any individual household in the following steps.

*Step* 1. Assign households in the data to one of the segments identified by our model based on each household's posterior segment probabilities and use segment-specific parameter estimates for each household.[11]

*Step* 2. Compute the objective function for a given household on a given shopping trip as the input for optimization. Specifically, compute $\hat{P}(I_{i,t}=1, B_{ik,t}=1)$ for previous-period buyers and previous-period nonbuyers separately, and compute $\hat{P}(I_{i,t+1}=1, B_{ik,t+1}=1)$ and $\hat{P}(I_{i,t+2}=1, B_{ik,t+2}=1)$ according to Equations (14) and (15) and Appendix B for the state-dependence parameter. Then, compute $E(Q_{ik,t+s} \mid I_{i,t+s}=1, B_{ik,t+s}=1)$, $s=0, 1, 2$, according to Equation (13). Finally, compute the objective function, Equation (12), using $\hat{P}(I_{i,t}=1, B_{ik,t}=1)$, $\hat{P}(I_{i,t+1}=1, B_{ik,t+1}=1)$, $\hat{P}(I_{i,t+2}=1, B_{ik,t+2}=1)$, and $E(Q_{ik,t+s} \mid I_{i,t+s}=1, B_{ik,t+s}=1)$, $s=0, 1, 2$. Keep $PC_{kt}$, $PC_{k,t+1}$, and $PC_{k,t+2}$ as the decision variables in the above computations.

*Step* 3. Derive the optimal price discounts $PC^*_{k,1}$, $PC^*_{k,2}$, and $PC^*_{k,3}$ *at the first shopping trip* during the

[11] An alternative approach is to follow the procedure we describe here for each segment and then take weighted average across segments to compute the objective function. We tested the predicted incidence and choice probabilities and purchase quantity in the holdout data using both approaches and found the difference to be negligible. We opted to use segment-specific estimates in the optimization because it reduces the amount of computation to about $1/G$ of the second approach (where $G$ is the number of segments in the model), which is desirable for real-world application.

time period of interest for a given household. Specifically, we use a constrained optimization routine in the software MATLAB® to maximize the objective function computed in Step 2 with respect to $PC_{k,1}$, $PC_{k,2}$, and $PC_{k,3}$, subject to the regularity constraints described previously. This yields the optimal price discounts $PC^*_{k,1}$, $PC^*_{k,2}$, and $PC^*_{k,3}$ for $t=1$.

*Step* 4. Derive the optimal price discounts *for the entire time period* of interest. Specifically, repeat the optimization procedure in Step 3 for each subsequent shopping trip $t$, $t=2,3,4,\ldots,$ using this household's purchase history data up to $t-1$. Steps 3 and 4 yield three sequences of optimal price discounts $\{PC^*_{k,t}\}$, $\{PC^*_{k,t+1}\}$, and $\{PC^*_{k,t+2}\}$, $t=1,2,3,\ldots,$ for the entire time period of interest. The final optimal promotion schedule over the time period is the first sequence $\{PC^*_{k,t}\}$. Note that $PC^*_{k,t+1}$ and $PC^*_{k,t+2}$ are auxiliary variables in the optimization procedure for each $t$, and they are updated by $PC^*_{k,t}$ at later shopping trips. For example, $PC^*_{k,t+1}$ derived at $t=1$ is replaced by $PC^*_{k,t}$ derived at $t=2$, and $PC^*_{k,t+2}$ derived at $t=1$ is replaced by $PC^*_{k,t}$ derived at $t=3$.

The above procedure can be performed on all households to get the customized promotion schedule for the entire sample of households in the data.

# 4. Application

## 4.1. Data Description

We apply our model and the promotion optimization method to purchase data provided by a leading online grocery retailer. The dataset contains a random sample of households from a midwest market over a 137-week period from 1997 to 1999. Two product categories, stick butter and liquid detergent, are included in the analysis. The first 48 weeks are used to initialize variables, the next 48 weeks are for model estimation, and the last 41 weeks are used as the holdout validation period.[12] Households that have made at least two purchases in the initialization period, two purchases in the estimation period, and one purchase in the holdout period are selected for analysis. The top four alternatives are included in the butter category: Land O'Lakes salted, Land O'Lakes unsalted, store brand salted, and store brand unsalted. They account for 95.3% of total category purchases and are all of the same size (four quarter-pound butter sticks per packet). We also selected the top four brands for liquid detergent: Wisk, All, Tide, and Cheer, which comprise 71.8% of total category purchases. Each brand combines various sizes. Table 1 presents descriptive statistics of the estimation data.[13] The brand distance measures were obtained from a survey (see Appendix A for details).

**Table 1 Descriptive Statistics of the Estimation Data**

| Alternative | No. of purchases | Average regular price (cents/oz.) | Average price cut (cents/oz.) |
|---|---|---|---|
| Butter (2,203 shopping trips, 129 households, and 48 weeks) | | | |
| Land O'Lakes Salted | 211 (26.5%) | 22.7 | 1.1/4.2 |
| Land O'Lakes Unsalted | 88 (11.1%) | 22.7 | 1.2/4.3 |
| Store brand, salted | 403 (50.6%) | 19.9 | 1.9/3.2 |
| Store brand, unsalted | 94 (11.8%) | 19.9 | 1.9/3.2 |
| Liquid laundry detergent (1,995 shopping trips, 108 households, and 48 weeks) | | | |
| Wisk | 117 (15.3%) | 7.1 | 0.77/0.84 |
| All | 53 (6.9%) | 4.6 | 0.30/0.48 |
| Tide | 538 (70.4%) | 7.0 | 0.47/0.88 |
| Cheer | 56 (7.3%) | 6.7 | 0.17/0.89 |

*Note.* For average price cut, the first number is the average over all weeks. The second number is the average over weeks when there was a price discount.

## 4.2. Parameter Estimates

We use Bawa's (1990) hybrid model as the benchmark to test our time-varying variety-seeking/inertia model. Note that the original hybrid model was a brand choice model, did not include marketing-mix variables, and was estimated on individual household data (see Bawa 1990). To compare the models on an equal ground, we extend the hybrid model such that the only difference is the formulation of the time-varying component of variety-seeking/inertia.[14] As described in §2, we employ the finite mixture model approach to handling heterogeneity. We allow all parameters in the brand utility function, category threshold function, and purchase quantity function to differ across segments.[15] The number of latent segments is determined based on the BIC. We find that a two-segment model is the best for both models in both categories. (In the butter category, the BIC is 2,855.4, 2,842.1, and 2,889.0 for the one-, two- and three-segment models, respectively, for our model, and the numbers are 3,009.1, 2,989.9, and 3,006.7 for

[12] The number of purchases per household during the estimation and holdout period ranges from 3 to 48 with a mean of 12.0 for butter, and ranges from 3 to 66 with a mean of 13.5 for liquid detergent.

[13] A concern in using this online shopping data set to estimate our model is that online consumers may have shopped from other outlets and those purchases were not recorded. To examine this issue, we compared purchase frequencies in our online data to those in brick-and-mortar store purchase data provided by ACNielsen. Our online consumers had slightly higher average purchase frequencies in both categories. Therefore, we are confident that these consumers shopped from the online store almost exclusively for the categories studied here.

[14] We thank an anonymous reviewer for suggesting this test.

[15] The model with segment-specific $\theta$, the interdependence parameter, does not offer significant improvement over the one with a common $\theta$ and, therefore, the latter is presented.

**Table 2 Parameter Estimates of Our Model**

| | Butter | | Liquid detergent | |
|---|---|---|---|---|
| Variables/parameter | Segment 1 | Segment 2 | Segment 1 | Segment 2 |
| Brand utility | | | | |
| Land O'Lakes, salted | −0.152 | 2.488*** | | |
| Land O'Lakes, unsalted | −0.135 | 2.193*** | | |
| Store brand, salted | −0.007 | 2.919*** | | |
| (Store brand, unsalted) | (0) | (0) | | |
| Wisk | | | 0.362 | 1.234** |
| All | | | −0.349 | −1.156** |
| Tide | | | 0.759*** | 1.728*** |
| (Cheer) | | | (0) | (0) |
| Regular price | −0.494*** | −0.036 | −0.307** | −0.825*** |
| Price cut | 0.676*** | 1.165** | 0.264 | 0.295** |
| $\gamma_0$ | 1.841*** | 1.101*** | 1.803*** | 1.169*** |
| $L_1$ | 0.356*** | 0.027 | 0.309*** | 0.297** |
| $L_1^2$ | −0.017** | −0.009 | −0.015* | −0.027** |
| $L_2$ | −1.381*** | −1.537*** | −1.826*** | −0.689** |
| $L_2^2$ | 1.124*** | 1.243*** | 4.870*** | 0.011 |
| $\varpi$ | 0.107 | −0.268** | −0.255** | 0.120 |
| Category threshold | | | | |
| Constant | 2.364*** | 4.428*** | 1.453*** | −1.918** |
| Purchase frequency | −4.501*** | −1.644** | −1.225* | −1.817** |
| Last purchase volume | 0.642** | 2.129** | 0.049* | 0.034** |
| Purchase quantity | | | | |
| Constant | −0.310 | −0.572*** | 5.849*** | 12.792*** |
| Regular price | −0.793*** | 0.074 | −0.569 | −14.102*** |
| Price cut | 0.853** | 1.043** | −0.073 | 2.754*** |
| Average purchase volume | 0.884* | 1.217** | 0.457*** | 0.955*** |
| Purchase frequency | | | −0.484 | −6.085*** |
| Segment size | 62.4% | 37.6% | 60.2% | 39.8% |
| Correlation ($\rho$) | −0.304 | | −0.251 | |
| $-LL$ | 2,695.9 | | 3,985.1 | |
| No. of parameters | 38 | | 40 | |

*Note.* Purchase frequency is insignificant in the purchase quantity equation in both segments and, thus, is dropped from the final model for the butter category.
* $P$-value $< 0.10$.
** $P$-value $< 0.05$.
*** $P$-value $< 0.01$.

**Table 3 Parameter Estimates of the Extended Bawa Hybrid Model**

| | Butter | | Liquid detergent | |
|---|---|---|---|---|
| Variables/parameter | Segment 1 | Segment 2 | Segment 1 | Segment 2 |
| Brand utility | | | | |
| Land O'Lakes, salted | 0.682*** | 1.301*** | | |
| Land O'Lakes, unsalted | 0.548** | −0.609* | | |
| Store brand, salted | 0.169*** | 4.127*** | | |
| (Store brand, unsalted) | (0) | (0) | | |
| Wisk | | | 0.381** | 0.847** |
| All | | | −1.540*** | −0.237 |
| Tide | | | 1.058*** | 5.192*** |
| (Cheer) | | | (0) | (0) |
| Regular price | −0.406*** | 0.196 | −0.760*** | −0.670* |
| Price cut | 0.476** | 1.471*** | 0.322*** | 0.177* |
| $b$ | 0.551*** | 1.690*** | 0.476*** | 1.930*** |
| $c$ | −0.018*** | −0.568*** | −0.013*** | −0.583*** |
| Category threshold | | | | |
| Constant | 2.701*** | 5.889*** | −1.763** | 1.466*** |
| Purchase frequency | −3.682*** | −1.933** | −0.090 | −2.992*** |
| Last purchase volume | 0.104 | 0.378* | 0.039*** | 0.026** |
| Purchase quantity | | | | |
| Constant | −0.626*** | −0.106 | 5.921*** | 8.074*** |
| Regular price | −0.050 | −1.043*** | −2.667*** | −7.260*** |
| Price cut | 0.399** | 1.484*** | 0.989* | 1.610*** |
| Average purchase volume | 1.419** | 0.099 | 0.561*** | 0.946*** |
| Purchase frequency | | | −2.570*** | −6.749*** |
| Segment size | 65.7% | 34.3% | 64.3% | 35.7% |
| Correlation ($\rho$) | −0.304 | | −0.260 | |
| $-LL$ | 2,874.4 | | 4,126.22 | |
| No. of parameters | 30 | | 32 | |

*Note.* Parameters b and c are coefficients of the linear and quadratic terms of the number of times an alternative has been purchased consecutively in the original Bawa model (1990). They are used to depict time-varying patterns of variety-seeking/inertia.
* $P$-value $< 0.10$.
** $P$-value $< 0.05$.
*** $P$-value $< 0.01$.

the extended Bawa model. In the detergent category, the numbers are 4,364.4, 4,137.1, and 4,175.3 for our model, and 4,512.7, 4,247.8, and 4,262.7 for the extended Bawa model.) The BIC measure also indicates that our model provides better fit to data than the hybrid model.

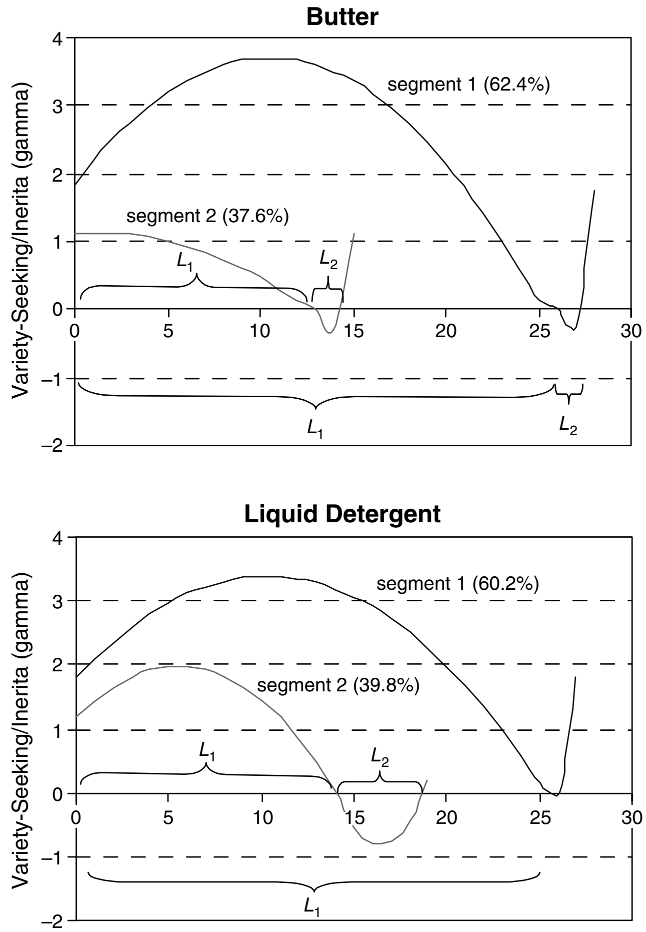
We present the parameter estimates of our model in Table 2 and the extended Bawa hybrid model in Table 3.[16] The common components in the two models are the category threshold utility function (Equation (2)), the latent purchase quantity function (Equation (6)), and the constants and price/promotion parameters in the brand utility functions in Equations (9a) and (9b). The parameters representing time-varying patterns of variety-seeking/inertia in the original Bawa model ($b$ and $c$) are of the expected signs. Parameter estimates of both models in both categories indicate that the effects of marketing-mix and household-specific variables in the three purchase decisions are also of the expected direction. Specifically, regular price negatively affects an alternative's choice probability as well as the purchase incidence probability of the category, while the reverse holds for price discount. A higher household purchase frequency (in the initialization period) is associated with a lower category incidence threshold and thus more frequent purchases, and the quantity bought on the last purchase occasion increases the threshold on the current shopping trip and thus reduces the purchase incidence probability. In addition, purchase quantity decreases with regular price

[16] We find that setting the quantity scale parameter $\delta_\xi$ to 1 provides more stable parameter estimates and better prediction. This was also done by Arora et al. 1998 (p. 39).

and increases with price cut of an alternative, and a household's average purchase quantity (and purchase frequency for the detergent category) in the initialization period are positively associated with the quantity purchased on a given occasion.

Our model can accommodate a wide variety of ways by which variety-seeking/inertia may change over time. Based on behavioral rationale, however, we expect both inertia and variety-seeking to first increase and then decline over time. In addition, we also expect price promotions to have a negative effect on purchase event feedback (Gedenk and Neslin 1999). As described in §2, the expected signs of the purchase event feedback effect parameters are $\omega_i$ negative, $\mu_{1i}$ positive, $\eta_{1i}$ negative, $\mu_{2i}$ negative, and $\eta_{2i}$ positive. The expected signs are supported by all *significant* parameter estimates in both categories (see Table 2). The parameter estimates also indicate that the more inertial segment is also less price and promotion sensitive in general. We illustrate the time-varying patterns of the variety-seeking/inertia parameter $\gamma_t$ in Figure 1. For each segment, $\gamma_t$ is a function of variable $L_{1,t_m-1}$ or $L_{2,t_m-1}$ (see Equations (10a) and (10b)). As the graphs indicate, $\gamma_t$ does exhibit cyclical patterns in the data, which means that *variety-seeking/inertia tendency not only changes over time for a substantial portion of households in the data, but also in a cyclical pattern as we expected*. This finding supports our view that, in addition to targeting, timing should be treated as an important component of customized promotion decisions.

**Figure 1** Time Varying Patterns of Variety-Seeking and Inertia




*Note.* $L_1 = L_{1,t_m-1}$ and $L_2 = L_{2,t_m-1}$. One and only one of them is greater than zero by definition. The variety-seeking/inertia parameter $\gamma_t$ is a function of $L_1$ or $L_2$ (Equations (10a) and (10b)): (1) $\gamma_t = \gamma_0 + \mu_1 L_1 + \eta_1 L_1^2$ when $L_1 > 0$ and $L_2 = 0$; (2) $\gamma_t = \mu_2 L_2 + \eta_2 L_2^2$ when $L_2 > 0$ and $L_1 = 0$. Each curve in the graph reflects the two functions for a segment, where the part above zero is based on the first function, and the part below zero is based on the second function. Note that $L_2$ starts from zero after the curve reaches zero and $L_1 = 0$.

### 4.3. Validation of Our Model vs. Bawa's Hybrid Model

We use several measures to compare the predictive power of our model versus the extended Bawa (1990) model. One measure is the log-likelihood in the holdout data. For the butter category, $-LL$ of our model is 2,238.8 and that of the Bawa model is 2,410.1 ($N = 1{,}582$). For the liquid detergent category, $-LL$ of our model is 3,494.7 and that of the Bawa model is 3,614.7 ($N = 1{,}462$). We also compare hit rates of purchase incidence and brand choice, as well as the root mean squared error (RMSE) for the predicted purchase quantity on each shopping trip in the holdout data. To compute these quantities, we first obtain the segment-specific predicted purchase incidence probability, brand choice probabilities, and purchase quantity, and then get the predicted values for each household by averaging the segment-specific quantities weighted by the household's posterior segment probabilities. In the butter category, the hit rate of purchase incidence is 77.0% for our model and 77.9% for the Bawa model, the hit rate of brand choice given purchase incidence is 80.5% for our model and 74.6% for the Bawa model, and the RMSE for purchase quantity is 41.86 for our model and 43.02 for the Bawa model. In the detergent category, the hit rate of purchase incidence is 84.5% for our model and 66.0% for the Bawa model, the hit rate of brand choice given purchase incidence is 90.0% for our model and 77.1% for the Bawa model, and the RMSE of purchase quantity is 79.02 for our model and 83.18 for the Bawa model. Our model provides better predictive power than the Bawa model on all the measures in both categories except the hit rate of purchase incidence for butter. Consistent with our expectation, the biggest improvement of our model over the Bawa model is in the brand choice component, as the other two components are formulated the same in both models. Our model also yields moderate improvement in predicting purchase quantity (and purchase incidence for one category) due to the interdependence of the three purchase decisions. In conclusion, in addition to the conceptual appeal over the hybrid model as described in §2, our model also offers better fit to the data and better out-of-sample predictive power.

### 4.4. Implementing Customized Promotions

The optimization is applied to the butter data in the holdout period which included 41 weeks and 1,582 shopping trips. We use one brand to illustrate the implementation. To protect confidentiality, we call it "the target brand" and denote it by subscript $k$. The operation would be the same for all brands. We assume that the regular price and price discount for competitive brands are as they appeared in the data.[17] Because $M_{kt}$, the manufacturer's regular profit margin without a promotion, is unavailable in our data, we assume three levels of the wholesale margin $m_m$ (20%, 30%, and 40%) and three levels of the retail margin $m_r$ (20%, 30%, and 40%) and examine the resulting nine scenarios.[18] $b_k$, the fraction of regular price for a maximum price cut, is set to 0.5 to reflect the deepest discount in the data ("buy one get one free").

We derive the optimal price discount schedule $\{PC^*_{k,t}\}$ for two types of prior purchase outcomes (previous-period buyers and previous-period nonbuyers). For both cases, we compute the expected category purchase incidence probability, brand choice probability, purchase quantity, expected cost of price promotion, and expected profit. We also compare our optimization approach to the following benchmarks.

**Benchmark 1.** Optimization based on the model with constant state dependence ($\gamma$) and the effect of last brand being purchased on promotion ($\omega$). This model allows current purchase to affect future purchase but does not take into account the time-varying patterns of $\gamma$. The optimal price promotion varies over time in this approach.

**Benchmark 2.** Optimization based on the model without purchase event feedback (no $\gamma$ and $\omega$).[19] In this model, there is no dynamics in the purchase behavior and, thus, the objective function only needs to include one period. The optimal price promotion may still vary over time depending on whether competitors change their prices and promotions.

**Benchmark 3.** No promotion at all.

**Benchmark 4.** A heuristic of customized promotions that approximates how Catalina Marketing Corporation distributes competitive coupons and loyalty coupons by its Checkout Coupon® service.[20] Specifically, for previous-period nonbuyers (which resemble competitive coupons), the average amount of price discount during a three-month bracket is offered when a household purchased a competitive brand at the previous shopping trip; for previous-period buyers (which resemble loyalty coupons), the average amount of price discount during a three-month bracket is offered when a household purchased the target brand at the previous shopping trip. The amount of price discount, if offered, does not vary in the short run according to this heuristic.

**Benchmark 5.** Actual practice in the holdout data. Table 4 presents the results. The purchase incidence and brand choice probabilities are averaged across all observations in the holdout data. The expected quantity, cost of promotion, and profit are summed over all observations in the data and measured in dollars. We also report the range of the optimal price discount according to each method. Table 4 indicates that our optimization method would outperform the five benchmarks, especially under the scenarios where promotion is more profitable. Under the lowest-profit scenario considered ($m_m = 20\%$, $m_r = 40\%$), our optimization and the two model-based methods (Benchmarks 1 and 2) all point to no promotion as the optimal strategy, because the expected cost of promotion under each method is zero or virtually zero. A closer look at the expected incidence, choice probabilities, and purchase quantity reveal that our optimization method, as well as Benchmarks 1–3, achieve higher profit than Benchmark 4 and the actual practice mainly through substantially cutting down unnecessary promotions for the category analyzed. Although the average purchase incidence probability, brand choice probability, and quantity are lower under the first four approaches, savings from costly promotions make up for the slight drop in expected quantity and renders higher expected profit than Benchmark 4 and actual practice. This also explains why Benchmark 3, which calls for no promotions, performs very well for the target butter brand. Table 4 also illustrates the benefit of performing optimization based on our model versus a model that assumes constant state dependence

[17] This is reasonable for customized promotion programs that are sponsored by a single manufacturer in a given product category, as we propose in this research. The retailer is free to accept other forms of trade promotions from other manufacturers as usual. In practice, those "regular" promotions are scheduled into a calendar in advance and are unlikely to deviate from the plan during the coverage of the promotion calendar.

[18] The manufacturer's profit and optimal price discount depend on $m_m(1 - m_r)$, with profit increases with $m_m(1 - m_r)$ for any given price discount. Due to space limitations, we present the three scenarios with the lowest, middle, and highest $m_m(1 - m_r)$. The complete results on nine scenarios are available upon request. In addition, the most likely scenario is the middle one with $m_m = 30\%$ and $m_r = 30\%$, based on information provided by industry sources.

[19] The best fitting model has three latent segments in this case.

[20] Inferred based on information from Catalina's company web site (www.catmktg.com). The exact algorithms are proprietary.

**Table 4 Optimization Result and Comparison to Benchmarks (Butter Data)**

| | $m_m$ (%) | $m_r$ (%) | Range of price cut ($/lb.) | Incidence probability | Choice probability | Expected quantity (lb) | Expected cost ($) | Expected profit ($) |
|---|---|---|---|---|---|---|---|---|
| Previous-period nonbuyers ($N = 1{,}096$) | | | | | | | | |
| Optimal | 20 | 40 | [0, 0.50] | 0.417 | 0.016 | 132.87 | 0.01 | 71.76 |
| | 30 | 30 | [0, 0.55] | 0.418 | 0.018 | 150.56 | 17.86 | 124.48 |
| | 40 | 20 | [0, 1.29] | 0.423 | 0.030 | 241.89 | 144.74 | 203.83 |
| Benchmark 1 | 20 | 40 | [0, 0.04] | 0.417 | 0.016 | 130.16 | 0.01 | 70.29 |
| | 30 | 30 | [0, 0.60] | 0.419 | 0.020 | 162.80 | 38.64 | 115.24 |
| | 40 | 20 | [0.18, 1.19] | 0.424 | 0.032 | 259.59 | 190.83 | 183.18 |
| Benchmark 2 | 20 | 40 | [0, 0.50] | 0.417 | 0.017 | 132.81 | 3.43 | 68.32 |
| | 30 | 30 | [0, 0.50] | 0.419 | 0.020 | 158.74 | 31.56 | 118.49 |
| | 40 | 20 | [0, 0.88] | 0.422 | 0.028 | 226.71 | 133.41 | 193.13 |
| Benchmark 3 | 20 | 40 | [0, 0] | 0.417 | 0.016 | 130.15 | 0 | 70.25 |
| | 30 | 30 | | | Same | | | 122.97 |
| | 40 | 20 | | | | | | 187.42 |
| Benchmark 4 | 20 | 40 | [0, 1.80] | 0.424 | 0.033 | 263.06 | 139.35 | −111.50 |
| | 30 | 30 | | | Same | | | −5.07 |
| | 40 | 20 | | | | | | 125.01 |
| Actual practice | 20 | 40 | [0, 2.25] | 0.424 | 0.035 | 280.00 | 291.43 | −140.40 |
| | 30 | 30 | | | Same | | | −27.13 |
| | 40 | 20 | | | | | | 111.32 |
| Previous-period buyers ($N = 486$) | | | | | | | | |
| Optimal | 20 | 40 | [0, 0] | 0.273 | 0.181 | 435.14 | 0 | 234.72 |
| | 30 | 30 | [0, 0.34] | 0.279 | 0.192 | 465.01 | 16.73 | 422.29 |
| | 40 | 20 | [0, 0.93] | 0.302 | 0.225 | 511.20 | 89.45 | 646.17 |
| Benchmark 1 | 20 | 40 | [0, 0] | 0.273 | 0.181 | 435.14 | 0 | 234.72 |
| | 30 | 30 | [0, 0.31] | 0.275 | 0.184 | 442.01 | 13.63 | 403.60 |
| | 40 | 20 | [0, 0.82] | 0.294 | 0.210 | 496.53 | 126.19 | 588.20 |
| Benchmark 2 | 20 | 40 | [0, 0] | 0.273 | 0.181 | 435.14 | 0 | 234.72 |
| | 30 | 30 | [0, 0.36] | 0.280 | 0.192 | 456.28 | 31.83 | 398.91 |
| | 40 | 20 | [0, 0.88] | 0.296 | 0.215 | 504.35 | 126.14 | 599.50 |
| Benchmark 3 | 20 | 40 | [0, 0] | 0.273 | 0.181 | 435.14 | 0 | 234.72 |
| | 30 | 30 | | | Same | | | 410.75 |
| | 40 | 20 | | | | | | 625.91 |
| Benchmark 4 | 20 | 40 | [0, 1.80] | 0.327 | 0.253 | 586.75 | 407.26 | −131.03 |
| | 30 | 30 | | | Same | | | 106.28 |
| | 40 | 20 | | | | | | 396.31 |
| Actual practice | 20 | 40 | [0, 2.25] | 0.336 | 0.269 | 623.59 | 584.54 | −248.29 |
| | 30 | 30 | | | Same | | | 3.90 |
| | 40 | 20 | | | | | | 312.14 |

*Notes.* Probabilities are averaged across shopping trips and expected quantity, cost, and profit are sums in the data. Expected Cost refers to expected cost of price promotion. For each observation in the data, it is computed as the multiplication of the unconditional choice probability of the target brand and its price cut.

(Benchmark 1) or one that does not account for purchase event feedback (Benchmark 2), and shows that the advantage of our approach over alternative ones increases with the manufacturer's profit margin. Interestingly, Benchmark 2 performs slightly better than Benchmark 1 under some scenarios in this analysis. It is likely due to better capturing of heterogeneity through a three-segment model without state dependence as in Benchmark 2 than a two-segment model with *constant* state dependence as in Benchmark 1.[21] Finally, comparing results for the two types of prior purchase outcomes, the expected cost of price promotion is lower and the expected profit is higher for previous-period buyers than for previous-period nonbuyers under our optimization method.

To further examine problems with the actual price promotions in the data, we classify the observations in the validation data into the following three groups: "agree" are those cases in which the actual price discount falls within 0.31 cents/oz. (i.e., five cents per packet) of the recommended optimal level, "missed opportunities" are those in which the actual price discount is more than 0.31 cents/oz. lower than the optimal level, and "wasted promotions" are

[21] We thank the area editor for this insight.

**Table 5 Types of Shopping Trips in the Validation Data**

| Scenario | | | | |
|---|---|---|---|---|
| $m_m$ (%) | $m_r$ (%) | Agree[(a)] | Missed opportunities[(b)] | Wasted promotions[(c)] |
| Previous-period nonbuyers ($N = 1{,}096$) | | | | |
| 20 | 40 | 454 (41.4%) | 19 (1.7%) | 623 (56.8%) |
| 30 | 30 | 242 (22.1%) | 231 (21.1%) | 623 (56.8%) |
| 40 | 20 | 204 (18.6%) | 368 (33.6%) | 524 (47.8%) |
| Previous-period buyers ($N = 486$) | | | | |
| 20 | 40 | 201 (41.4%) | 0 | 285 (58.6%) |
| 30 | 30 | 148 (30.5%) | 53 (10.9%) | 285 (58.6%) |
| 40 | 20 | 159 (32.7%) | 74 (15.2%) | 253 (52.1%) |

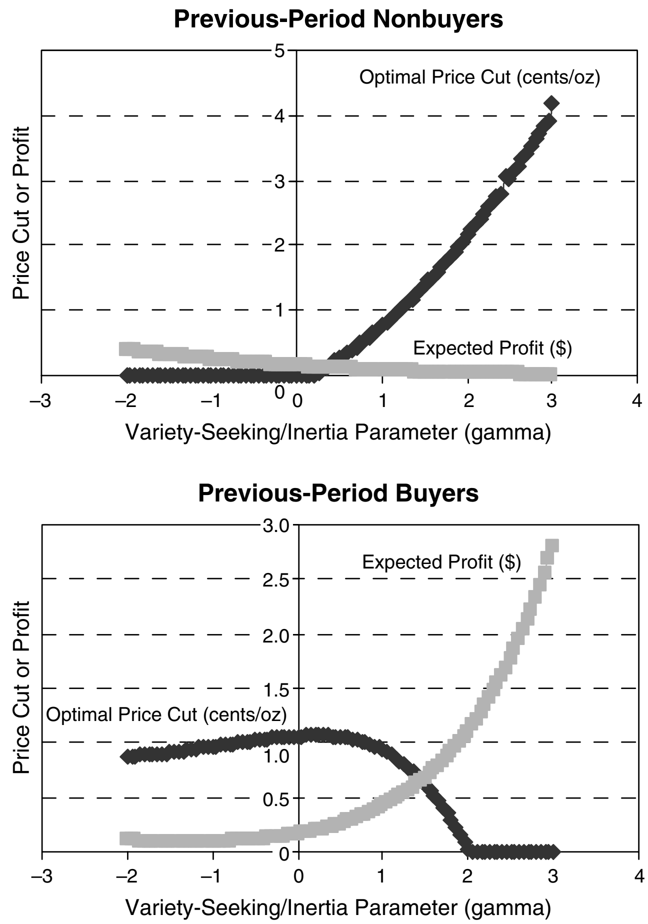
[(a)] Agree: An actual price discount falls within 0.31 cents/oz. (i.e., 5 cents per packet) of the recommended optimal level.
[(b)] Missed opportunity: An actual price discount is more than 0.31 cents/oz. lower than the recommended optimal level.
[(c)] Wasted promotion: An actual price discount is more than 0.31 cents/oz. higher than the recommended optimal level.

those in which the actual price discount is more than 0.31 cents/oz. higher than the optimal level. Table 5 summarizes the number in each category for the two types of prior purchase outcomes under three profit scenarios. It appears that wasted promotions were a serious problem, accounting for more than 50% of the observations. This is consistent with the message revealed in Table 4. It suggests that the manufacturer could have reduced substantial promotions and better allocated the spending by shifting from wasted promotions to missed opportunities.

### 4.5. Timing Promotions by Variety-Seeking/Inertia Tendency

The optimal price discount on each shopping trip is influenced by external factors such as competitive prices and promotions, as well as consumers' internal factors such as brand preferences and variety-seeking/inertia tendency. Our model estimation results show that the tendency does vary over time for most households. Therefore, it is important to adjust promotions by tracking each household's variety-seeking/inertia over time. To examine the relationship between variety-seeking/inertia and the optimal price cut and expected profit, one needs to control for the other variables in the objective function. Figure 2 presents a simulation result which assumes the most likely scenario of $m_m = 30\%$ and $m_r = 30\%$ with parameters taking the values obtained in the bigger segment. The continuous variables in the objective function are set at their average values and the categorical variables are assumed equal to their mode. Note that all parameters and variables, except the variety-seeking/inertia parameter $\gamma$, are set to constant values. In the simulation, 101 data points are generated with $\gamma$ taking value from $-2$ to 3 with a step-size of 0.05.

**Figure 2 Optimal Price Cut and Expected Profit vs. the Variety-Seeking/Inertia Parameter**




The upper graph in Figure 2 is for previous-period nonbuyers. It shows that the optimal price cut increases with $\gamma$ while the expected profit decreases with $\gamma$. In fact, no price cut is needed when $\gamma$ is below 0.2, i.e., when a consumer is mainly in variety-seeking mood. The lower graph in Figure 2 is for previous-period buyers. The optimal price cut increases first when $\gamma$ is in the high variety-seeking range and then drops as $\gamma$ increases. No price cut would be needed if a consumer's inertia tendency is high enough ($\gamma \geq 2.0$ in this case). The expected profit increases monotonically with the inertial tendency in this case. These patterns imply that a manager should pay close attention to the time-varying pattern of each household's variety-seeking/inertia tendency when making customized promotion decisions, and that the optimal price cut and expected profit differ vastly depending on whether the brand was chosen on the previous purchase occasion.

### 4.6. Validation of Our Customization Method

To examine how well our customization method would work in reality, we compare the *actual profit* in

**Table 6** $t$**-Tests of Actual Profit in the Validation Data**

| Scenario | | | | |
|---|---|---|---|---|
| $m_m$ (%) | $m_r$ (%) | Agree (cents) | Missed opportunities and wasted promotions (cents) | P-value |
| Previous-period nonbuyers | | | | |
| 20 | 40 | 2.2 | −5.5 | <0.0001 |
| 30 | 30 | 3.9 | −1.0 | 0.0053 |
| 40 | 20 | 6.6 | 2.1 | <0.0001 |
| Previous-period buyers | | | | |
| 20 | 40 | 18.8 | −36.6 | <0.0001 |
| 30 | 30 | 35.7 | −10.1 | <0.0001 |
| 40 | 20 | 54.1 | 11.3 | <0.0001 |

*Note.* Average profit on each shopping trip is measured as cents.

the holdout data of those shopping trips which happened to comply with the definition of agree to those of missed opportunities and wasted promotions. The latter two categories are grouped together and $t$-tests are conducted on the agree cases against the other two. Note that there is no a priori guarantee that the agree cases based on our method would render higher profit because the comparison is performed on the holdout data which were not used for model estimation. Table 6 summarizes the result. It is found that the actual average profit (measured as cents per shopping trip) on shopping trips agreeing with our recommendations was significantly higher than that of missed opportunities and wasted promotions for both types of prior purchase outcomes under all scenarios. The average actual profit in the agree cases were all positive, while it was negative under most scenarios (including the most likely scenario) in the missed opportunities and wasted promotions cases.[22] In addition, the customization method is especially effective for previous-period buyers in terms of the increment in profit. In summary, our proposed customization method can greatly improve the effectiveness of the online retailer's current promotion practice by directing promotion spending to purchase opportunities that would yield profit and avoiding unnecessary spending in other cases. The improvement is especially substantial when aimed at consumers who purchased the target brand on the previous purchase occasion.

## 5. Discussion

The rapid growth of Internet technology has stimulated a growing body of academic research.[23] In our view, the Internet has provided a great opportunity to not only deliver personalized promotions to individual households, but also update the offerings based on each household's most recent purchase data. This research is an attempt to develop an actionable promotion customization system that takes advantage of the one-on-one and interactive nature of the online shopping environment. It is built upon a model that allows variety-seeking/inertia to vary over time for the same household and derives the optimal price discount for each household on each shopping trip by taking into account the impact of current promotions on future purchases.

Our model estimation results indicate that the variety-seeking/inertia tendency does vary over time in a cyclical pattern for most households. This supports our view that *timing*, in addition to targeting, should be treated as an important component of customized promotion decisions when the shopping environment, such as the Internet, provides the opportunity to update offerings on each shopping trip. The importance of the timing issue is further substantiated by the comparison of our optimization approach and the ones based on a model with constant variety-seeking/inertia and one that does not take into account the purchase event feedback effect. Our approach outperforms the alternatives under all scenarios analyzed and its advantage becomes greater under the high-profit scenarios. Interestingly, we find that, in the butter category, there appears to be too much promotion, and a heuristic which calls for no promotion could substantially increase the expected profit of the target brand. This may be a simple guideline for butter manufacturers in real practice, although our approach could help achieve even higher profit. In addition, we find that the proposed method would be especially effective when offered to consumers who purchased the targeted brand on the previous purchase occasions.

Our model and customization approach are oriented toward consumer packaged goods. The model applies to consumer products that have clearly defined category and choice alternatives and are bought frequently. For nonpackaged goods such as books, music CDs, and apparels, consumers do not buy the exactly same alternative repeatedly in general, so the operationalization of variety-seeking/inertia as in our model would not apply. Also, for these products, the decision on what items to promote to which consumer is more important than when and how much to discount for a given item out of a fixed set. For retailers like Amazon, CD Now, and Lands End, a product recommendation system (such as the one proposed by Ansari et al. 2000) would be more relevant.

A limitation of our consumer response model is that it does not take into account consumers' expectations of future promotions when making purchase

[22] The specific profit in our analysis depends substantially on the values assumed for $m_m$ and $m_r$. The current promotion policy could be more profitable than our numbers may suggest, because we do not have access to trade deals or other cost data. We thank the area editor for raising this point.

[23] See Park and Fader (2003), Chatterjee et al. (2003), and Danaher et al. (2003) for some recent examples.

decisions. It has been documented that many consumers hold beliefs about future promotion offerings and may adjust their purchase behavior accordingly (e.g., Krishna 1992, Gönül and Srinivasan 1996, Erdem et al. 2003, Sun et al. 2003). Utilizing a dynamic structural model of purchase incidence and brand choice, Sun et al. (2003) show that ignoring this consumer forward-looking behavior can lead to biased estimates of brand switching and price coefficients. Therefore, caution should be taken when applying our model and optimization approach. Recently, Erdem et al. (2003) and Hendel and Nevo (2002) have developed dynamic models of purchase incidence, brand choice, and purchase quantity that accommodate consumers' expectation of future price changes. These state-of-the-art models are more demanding computationally. A justification for using our consumer response model is its relative simplicity, which is desirable for a decision-support system in real practice.

This research is mainly aimed at providing a customized promotion system for online stores. Nevertheless, the proposed method is not limited to cyberspace. The technical requirement for implementing our decision-support system is the ability to modify and update promotion offerings to individual households on each shopping trip. Currently, the Internet is the predominant venue that fits this requirement. It is worth noting, though, that some firms have been experimenting with interactive personalized advertising and promotions in brick-and-mortar stores (e.g., Klever Marketing, Inc.'s Klever-Kart® system[24]). Our customized promotion system can be utilized by those firms as well.

This study points to new business opportunities for online retailers as well as new promotion strategies for consumer packaged goods manufacturers. Online retailers can take advantage of their ability to deliver interactive and one-on-one marketing actions and offer customized promotion programs to manufacturers. These value-added services could bring in profit in addition to their regular businesses. Manufacturers of consumer packaged goods also benefit from the proposed customization services which could greatly improve their promotion effectiveness and efficiency. We have found that the proposed customization method would be especially effective when offered to consumers who purchased the targeted brand on previous purchase occasions. Because a dominant brand is more likely to be purchased on previous purchase occasions by consumers, while a small brand is less likely to be purchased on previous purchase occasions, we conjecture that the proposed customization method would be particularly attractive to manufacturers with large brands in the marketplace.

[24] The reader is referred to www.kleverkart.com for details.

**Acknowledgments**

The authors thank an anonymous online retailer for generously providing the data used in this study. We are grateful to Fred Feinberg, Sachin Gupta, Dipak Jain, Aradhna Krishna, Peter Popkowski Leszczyc, and Michel Wedel for their helpful input. Financial support from Procter & Gamble's Marketing Innovation Research Fund to the first author is greatly appreciated. We also thank the former and current editors, the area editor, and three anonymous reviewers for their many valuable comments and suggestions. This paper is based on the doctoral dissertation of the first author.

## Appendix A. Brand Distance Measures

Subjects were asked to rate the similarity/dissimilarity of pairs of choice alternatives on a 1–9 point scale. Three product attributes were provided for each choice alternative. For the butter category the attributes were brand, flavor, and regular price. For the liquid detergent category the attributes were brand, manufacturer, and regular price. The similarity/dissimilarity scale and order of product pairs were counterbalanced in the questionnaires. In computing the distance measures, similarity was first converted to dissimilarity on a 1–9 point scale. Then, the raw scores were rescaled for each subject such that the maximum distance is set to 1 and the rest is a ratio of the raw score to the maximum distance. This was done to control for difference in the subjects' anchoring point. Next, we obtained averages of the rescaled distance scores. Finally, to control for the scale of the variety-seeking/inertia parameter, the average scores were normalized such that the maximum was set to 1 and the rest is a ratio of the original score to the maximum. One hundred subjects participated in the study. The normalized brand distance measure ranges from 0.575 to 1 for butter and 0.564 to 1 for liquid detergent. (More details are available from the authors upon request.)

## Appendix B. Variety-Seeking/Inertia Parameter in the Future Utility Functions

For previous-period nonbuyers, let $k$ be the target brand, $l$ be the brand chosen at the previous purchase occasion prior to $t$, and $j$ be any brand other than $k$ and $l$. For previous-period buyers, let $k$ be the target brand, $j \neq k$ be any other brand, and $t$ be the shopping trip of interest.

$$\gamma_{t+1} = \begin{cases} \text{previous-period nonbuyers } (l \text{ was chosen at } t-1)\text{:} \\ \quad \begin{cases} \text{no purchase at } t\text{: } \gamma_t \\ l \text{ chosen at } t\text{: } \begin{cases} l \text{ with prom: } \gamma_t \\ l \text{ without prom:} \\ \qquad \gamma_t + \mu_1 + \eta_1 + 2L_{1,\,t-1} \end{cases} \\ k \text{ chosen at } t\text{: } \gamma_t \\ j \neq l, k \text{ chosen at } t\text{: } \begin{cases} j \text{ with prom: } \gamma_t \\ j \text{ without prom: } \mu_2 + \eta_2 \end{cases} \end{cases} \\ \text{previous-period buyers } (k \text{ was chosen at } t-1)\text{:} \\ \quad \begin{cases} \text{no purchase at } t\text{: } \gamma_t \\ k \text{ chosen at } t\text{: } \gamma_t \\ j \neq k \text{ chosen at } t\text{: } \begin{cases} j \text{ with prom: } \gamma_t \\ j \text{ without prom: } \mu_2 + \eta_2 \end{cases} \end{cases} \end{cases}$$

$$\gamma_{t+2} \quad \text{(same for the two types of buyers)}$$

$$= \begin{cases} \text{no purchase at } t+1\text{: } \gamma_t \\ k \text{ chosen at } t+1\text{: } \gamma_{t+1} \\ j \neq k \text{ chosen at } t+1 \text{ \& } l \neq j \text{ chosen at } t\text{:} \\ \quad \begin{cases} j \text{ with prom: } \gamma_{t+1} \\ j \text{ without prom: } \mu_2 + \eta_2 \end{cases} \\ j \neq k \text{ chosen at } t+1 \text{ \& } j \text{ chosen at } t\text{:} \\ \quad \begin{cases} j \text{ with prom: } \gamma_{t+1} \\ j \text{ without prom: } \gamma_{t+1} + \mu_1 + \eta_1 + 2L_{1,t} \end{cases} \end{cases}$$

## Appendix C. Derivation of the Joint Probability

Let

$$A = \ln \sum_{\substack{j=0,1,\ldots,K \\ \text{and } j\neq k}} e^{V_{ijt}}.$$

$$\begin{aligned}
&\Pr(I_{it}=1, B_{ikt}=1, Q_{ikt}=q_{ikt}) \\
&\quad = \int_{-\infty}^{V_{ikt}} f(\varepsilon^*_{ikt}, \xi_{ikt}=q_{ikt}-Z_{ikt}\phi_i)\, d\varepsilon^*_{ikt} \\
&\quad = \frac{\partial F(\varepsilon^*_{ikt}, \xi_{ikt})}{\partial \xi_{ikt}}\bigg|_{\substack{\varepsilon^*_{ikt}=V_{ikt} \\ \xi_{ikt}=q_{ikt}-Z_{ikt}\phi_i}} \\
&\quad = \frac{\partial}{\partial \xi_{ikt}}\left\{\frac{1}{1+e^{A-\varepsilon^*_{ikt}}}\frac{1}{1+e^{-\delta_\xi \xi_{ikt}}}\left[1+\theta\left(1-\frac{1}{1+e^{A-\varepsilon^*_{ikt}}}\right)\right.\right. \\
&\qquad\qquad \left.\left.\cdot\left(1-\frac{1}{1+e^{-\delta_\xi\xi_{ikt}}}\right)\right]\right\}\bigg|_{\substack{\varepsilon^*_{ikt}=V_{ikt} \\ \xi_{ikt}=q_{ikt}-Z_{ikt}\phi_i}} \\
&\quad = \frac{\delta_\xi e^{-\delta_\xi \xi_{ikt}}}{(1+e^{A-\varepsilon^*_{ikt}})(1+e^{-\delta_\xi\xi_{ikt}})^2} \\
&\qquad \cdot\left[1+\theta\frac{e^{A-\varepsilon^*_{ikt}}}{(1+e^{A-\varepsilon^*_{ikt}})}\frac{-1+e^{-\delta_\xi\xi_{ikt}}}{1+e^{-\delta_\xi\xi_{ikt}}}\right]\bigg|_{\substack{\varepsilon^*_{ikt}=V_{ikt} \\ \xi_{ikt}=q_{ikt}-Z_{ikt}\phi_i}} \\
&\quad = \frac{e^{V_{ikt}}}{\sum_{j=0}^K e^{V_{ijt}}}\frac{\delta_\xi e^{\delta_\xi(Z_{ikt}\phi_i-q_{ikt})}}{[1+e^{\delta_\xi(Z_{ikt}\phi_i-q_{ikt})}]^2} \\
&\qquad \cdot\left[1+\theta\left(1-e^{V_{ikt}}\Big/\sum_{j=0}^K e^{V_{ijt}}\right)\frac{-1+e^{\delta_\xi(Z_{ikt}\phi_i-q_{ikt})}}{1+e^{\delta_\xi(Z_{ikt}\phi_i-q_{ikt})}}\right].
\end{aligned}$$

## Appendix D. Derivation of the Conditional Expectation of Purchase Quantity

Let

$$B = Z_{ikt}\phi_i, \quad C = 1 - e^{V_{ikt}}\Big/\sum_{j=0}^K e^{V_{ijt}}.$$

$$\begin{aligned}
E(Q_{ikt}\mid I_{it}=1, B_{ikt}=1) &= \int_0^{+\infty}\Pr(Q_{ikt}=q\mid I_{it}=1,B_{ikt}=1)\,q\,dq \\
&= \int_0^{+\infty}\frac{\delta_\xi e^{\delta_\xi(B-q)}}{[1+e^{\delta_\xi(B-q)}]^2} \cdot\left[1+\theta C\frac{-1+e^{\delta_\xi(B-q)}}{1+e^{\delta_\xi(B-q)}}\right]q\,dq \\
&= \int_0^{+\infty}\frac{\delta_\xi e^{\delta_\xi(B-q)}}{[1+e^{\delta_\xi(B-q)}]^2}q\,dq \\
&\quad +\theta C\int_0^{+\infty}\frac{\delta_\xi e^{\delta_\xi(B-q)}(e^{\delta_\xi(B-q)}-1)}{[1+e^{\delta_\xi(B-q)}]^3}q\,dq \\
&= \left[\frac{q}{[1+e^{\delta_\xi(B-q)}]}-\frac{1}{\delta_\xi}\log(e^{\delta_\xi B}+e^{\delta_\xi q})\right]_0^{+\infty} \\
&\quad +\theta C\left[\frac{e^{\delta_\xi(B-q)}(\delta_\xi q-1)-1}{\delta_\xi[1+e^{\delta_\xi(B-q)}]^2}\right]_0^{+\infty} \\
&= \frac{1}{\delta_\xi}\left[\log(1+e^{\delta_\xi B})-\theta C\frac{e^{\delta_\xi B}}{1+e^{\delta_\xi B}}\right].
\end{aligned}$$

Therefore,

$$\begin{aligned}
&E(Q_{ikt}\mid I_{it}=1, B_{ikt}=1) \\
&\quad = \frac{1}{\delta_\xi}\left[\log(1+e^{\delta_\xi Z_{ikt}\phi_i}) - \theta\left(1-e^{V_{ikt}}\Big/\sum_{j=0}^K e^{V_{ijt}}\right)\frac{e^{\delta_\xi Z_{ikt}\phi_i}}{1+e^{\delta_\xi Z_{ikt}\phi_i}}\right].
\end{aligned}$$

## References

Ansari, Asim, Rajeev Kohli, Skander Essegaier. 2000. Internet recommendation systems. *J. Marketing Res.* **37**(August) 363–375.

Arora, Neeraj, Greg M. Allenby, James L. Ginter. 1998. A hierarchical Bayes model of primary and secondary demand. *Marketing Sci.* **17**(1) 29–44.

Bawa, Kapil. 1990. Modeling inertia and variety seeking tendencies in brand choice behavior. *Marketing Sci.* **9**(3) 263–278.

Bell, David R., Jeongwen Chiang, V. Padmanabhan. 1999. The decomposition of promotional response: An empirical generalization. *Marketing Sci.* **18**(4) 504–526.

Ben-Akiva, Moshe, Steven R. Lerman. 1985. *Discrete Choice Analysis: Theory and Application to Travel Demand*. MIT Press, Cambridge, MA.

Chatterjee, Patrali, Donna L. Hoffman, Thomas P. Novak. 2003. Modeling the clickstream: Implications for web-based advertising efforts. *Marketing Sci.* **22**(4) 520–541.

Chiang, Jeongwen. 1991. A simultaneous approach to the whether, what, and how much to buy questions. *Marketing Sci.* **10**(Fall) 297–315.

Chintagunta, Pradeep K. 1993. Investigating purchase incidence, brand choice and purchase quantity decisions of households. *Marketing Sci.* **12**(Spring) 184–208.

Chintagunta, Pradeep K., Sudeep Haldar. 1998. Investigating purchase timing behavior in two related product categories. *J. Marketing Res.* **35**(February) 43–53.

Danaher, Peter J., Isaac W. Wilson, Robert A. Davis. 2003. A comparison of online and offline consumer brand loyalty. *Marketing Sci.* **22**(4) 461–476.

Erdem, Tülin, Susumu Imai, Michael Keane. 2003. A model of consumer brand and quantity choice dynamics under price uncertainty. *Quant. Marketing Econom.* **1**(1) 5–64.

Gedenk, Karen, Scott A. Neslin. 1999. The role of retail promotion in determining future brand loyalty: Its effect on purchase event feedback. *J. Retailing* **75**(4) 433–459.

Gönül, Füsun, Kannan Srinivasan. 1996. Estimating the impact of consumer expectations of coupons on purchase behavior: A dynamic structural model. *Marketing Sci.* **15**(3) 262–279.

Gumbel, E. J. 1961. Bivariate logistic distributions. *J. Amer. Statist. Association* **56**(294) 335–349.

Hanemann, W. Michael. 1984. Discrete continuous models of consumer demand. *Econometrica* **52** 541–561.

Hendel, Igal, Aviv Nevo. 2002. Sales and consumer inventory. Working paper 9048, National Bureau of Economic Research, Cambridge, MA.

Jain, Dipak C., Naufel J. Vilcassim. 1991. Investigating household purchase timing decisions: A conditional hazard function approach. *Marketing Sci.* **10**(1) 1–23.

Jeuland, Abel P. 1979. Brand choice inertia as one aspect of the notion of brand loyalty. *Management Sci.* **25** 671–682.

Kahn, Barbara E., Manohar U. Kalwani, Donald G. Morrison. 1986. Measuring variety-seeking and reinforcement behavior using panel data. *J. Marketing Res.* **23**(May) 89–100.

Kamakura, Wagner A., Gary J. Russell. 1989. A probabilistic choice model for market segmentation and elasticity structure. *J. Marketing Res.* **26**(November) 379–390.

Krishna, Aradhna. 1992. The normative impact of consumer price expectations for multiple brands on consumer purchase behavior. *Marketing Sci.* **11**(3) 266–286.

Krishnamurthi, Lakshman, S. P. Raj. 1988. A model of brand choice and purchase quantity price sensitivities. *Marketing Sci.* **7**(1) 1–20.

Lattin, James M. 1987. A model of balanced choice behavior. *Marketing Sci.* **6**(1) 48–65.

McFadden, Daniel. 1978. Modeling the choice of residential location. A. Karlquist et al. eds. *Spatial Interaction Theory and Residential Location.* North Holland, Amsterdam, The Netherlands, 75–96.

Papatla, Purushottam, Lakshman Krishnamurthi. 1992. A probit model of choice dynamics. *Marketing Sci.* **11**(2) 189–206.

Park, Young-Hoon, Peter S. Fader. 2003. Modeling browsing behavior at multiple websites. *Marketing Sci.* **23**(3) 280–303.

Rossi, Peter E., Robert E. McCulloch, Greg M. Allenby. 1996. The value of purchase history data in target marketing. *Marketing Sci.* **15**(4) 321–340.

Seetharaman, P. B., Pradeep Chintagunta. 1998. A model of inertia and variety-seeking with marketing variables. *Internat. J. Res. Marketing* **15**(1) 1–17.

Seetharaman, P. B., Andrew Ainslie, Pradeep Chintagunta. 1999. Investigating household state dependence effects across categories. *J. Marketing Res.* **36**(November) 488–500.

Shaffer, Greg, Z. John Zhang. 1995. Competitive coupon targeting. *Marketing Sci.* **14**(4) 395–416.

Silva-Risso, Jorge M., Randolph E. Bucklin, Donald G. Morrison. 1999. A decision support system for planning manufacturers' sales promotion calendars. *Marketing Sci.* **18**(3) 274–300.

Sun, Baohong, Scott Neslin, Kannan Srinivasan. 2003. Measuring the impact of promotions on brand switching when consumers are forward looking. *J. Marketing Res.* **40**(November) 389–405.

Tellis, Gerald J., Fred S. Zufryden. 1995. Tackling the retailer decision maze: Which brands to discount, how much, when and why? *Marketing Sci.* **14**(3) 271–299.

Trivedi, Minakshi, Frank Bass, Ram C. Rao. 1994. A model of stochastic variety-seeking. *Marketing Sci.* **13**(3) 274–297.

Zhang, Jie. 1999. Investigating dynamic brand choice processes: A comparison of online and store shopping environments. Unpublished doctoral dissertation, Northwestern University, Evanston, IL.