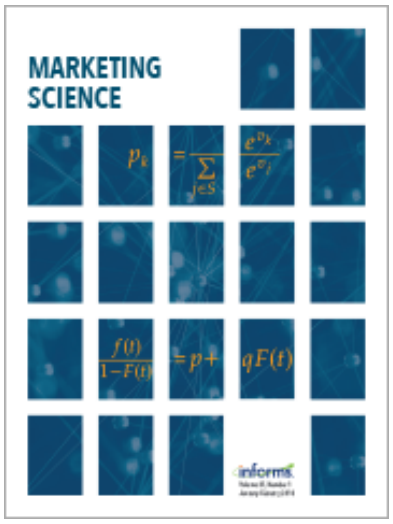


Marketing Science

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

Consumer Learning and Brand Valuation: An Application on Over-the-Counter Drugs

M. Tolga Akçura, Füsun F. Gönül, Elina Petrova,

To cite this article:
M. Tolga Akçura, Füsun F. Gönül, Elina Petrova, (2004) Consumer Learning and Brand Valuation: An Application on Over-the-Counter Drugs. Marketing Science 23(1):156-169. https://doi.org/10.1287/mksc.1030.0028

MARKETING SCIENCE
Vol. 23, No. 1, Winter 2004, pp. 156–169
ISSN 0732-2399 | EISSN 1526-548X | 04 | 2301 | 0156

informs

DOI 10.1287/mksc.1030.0028

Research Note

Consumer Learning and Brand Valuation: An Application on Over-the-Counter Drugs

M. Tolga Akçura
Krannert Graduate School of Management, Purdue University, West Lafayette, Indiana 47907-1310, akcurat@mgmt.purdue.edu

Füsun F. Gönül
Heinz School, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, www.fusunconsulting.com

Elina Petrova
Graduate School of Industrial Administration, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, epetrova@andrew.cmu.edu

We develop a brand choice model with learning based on the Kalman filter methodology. The model enables us to separate the effects of contemporaneous marketing promotions from the impact of the perceived quality valuation accrued through product usage over time. We also account for idiosyncratic consumer learning and preferences. The results point to the presence of heterogeneity in the valuation carryover coefficients across consumers and brands. In contrast to our expectations, a higher price is not important for most of the consumers in the sample. The model enables us to compare brands in terms of their memorability, which determines brand salience on the next purchase occasion. Our findings suggest that price promotions may be deficient as a tool to increase market share in the studied product category. The proposed model is applicable to other consumer goods contingent on consumers' being sufficiently motivated to learn their own preferences via personal experience. Brand managers can use the model for comparative diagnostics and market performance simulation under different price and promotion scenarios. This paper is instructive to the application of a relatively new methodology; we illustrate the analytical potential of the model by demonstrating its inferential power in a specific marketing context.

Key words: brand choice; buyer behavior; consumer learning; hierarchical Bayes analysis
History: This paper was received April 5, 2000, and was with the authors 24 months for 3 revisions; processed by Dick Wittink.

1. Introduction

Consider the following scenario: You suffer from muscle pain that is causing you some discomfort. You rummage through your medicine cabinet but find nothing appropriate. Your options are to see a physician or go to a drugstore. The first alternative requires that you call your physician's office and schedule an appointment, wait a few days before your visit, get a checkup and prescription (which, for all you know, might be for an over-the-counter (OTC) drug), pay for the portion of the checkup not covered by insurance, and go to the pharmacy to fill the prescription. The second alternative is considerably easier and faster—head for the closest pharmacy where you can promptly choose a drug from the pain relievers aisle.

The rapid growth of the OTC drugs market, which is expanding at a compounded annual rate of 7.5%, suggests that most people prefer the latter alternative despite the concerns voiced by many physicians about the risks of self-medication. This growth is partly due to OTC drugs' convenience and purchase immediacy, as well as consumers' feeling of control over the choice decision. However, there are costs associated with this decision (e.g., insurance plans do not cover most OTC drugs). Furthermore, consumers' decisions are often made with incomplete information about their preferences because they lack comparative or experiential knowledge, which might impose cognitive costs magnified by the perceived importance of the brand choice.

We suggest that each purchase of a specific OTC drug presents a learning opportunity for consumers to assess the effects of the drug on their health condition and on their physical well-being in general. This presumption is based on the mostly intermittent usage of such drugs. During the period between two treatments, a consumer's memory of the drug's efficacy might fade, the focal condition might alter, the application circumstances might change, or the product quality might be modified. As a result, a

consumer's valuations of a drug are likely to shift over time because of the dynamic acquisition of new information.

We set out to build a model that can disentangle the effect of evolving quality valuation from the effects of promotional efforts that affect the brand choice decision. To accomplish this, we propose a Kalman filter procedure, combined with hierarchical Bayes analysis. We model the underlying learning process as compounded by time-dependent forgetfulness that is manifested as a potential decline in the valuation carryover. We also examine how consumers develop individual product valuations through personal experience and whether updating and forgetting exhibit brand-specific effects. The decision to apply our model to the choice of OTC drugs is motivated by the following:

(i) The product category is characterized by a high level of personal involvement in the purchase decision. Consumers are aware that inadequate self-treatment can lead to potentially serious ramifications. Thus, consumers are expected to be sufficiently motivated to learn which drug is the best for their overall condition.

(ii) The perceived significance of making the right choice of a drug is enhanced by consumers' understanding that experiential learning will increase the chance for effective self-treatment in the future. The associated credibility of the experience makes the learning even more germane.

(iii) Taking medication is a highly personal experience. The same drug may have different effects on different people. It may even have a different impact on the same person when it is administered at different times and under different circumstances. OTC drugs present ample learning opportunities through consecutive treatments that are idiosyncratic among consumers.

(iv) The choice of an OTC drug gives consumers a sense of control when making decisions about their health, which is not the case with prescription drugs. Learning what serves their needs best may enhance consumers' perceptions of self-sufficiency and convenience.

(v) For forward-looking, rational consumers, learning their personal preferences and forming a judgment about the performance of OTC drugs can reduce future costs such as physical costs (e.g., unnecessarily extending the period of discomfort by choosing an incorrect drug), monetary costs (e.g., extra drug purchases, visits to physicians in case of suboptimal choice), and opportunity costs (e.g., time away from work, reduced quality of life).

(vi) The core benefit of most OTC drugs is relief from symptoms; thus, the immediate value of drug usage will dominate the future value of information obtained through experimentation with untried brands. We believe that in our case, brand-centered learning is likely to prevail over category learning, because the latter might be associated with greater risk and uncertainty.

2. Model

2.1. General Model Formulation Based on the Valuation Updating Process

We propose that before purchasing any particular brand in the leg-and-back pain reliever category, the consumer has a prior valuation of the brand's (expected) performance based on advertising, word of mouth, reference price, or general experience with other pain relievers. When taking the drug, the consumer observes the drug's impact on his or her condition, forms a judgment (perception) of the drug's efficacy, and changes his or her valuation of the drug accordingly. This process leads to the formation of a posterior valuation by the end of the period. It is this posterior valuation that determines a consumer's subsequent brand choice and serves as the new prior in the updating process (Figure 1). The mean value of the posterior valuation may be higher or lower than that of the prior one, depending on the latest perception of the drug efficacy and the current marketing activities of the brand.

A Kalman filter model is a two-equation model that accommodates the coexistence of two random processes, one in which updating occurs (to capture consumer learning) and the other in which no updating takes place (for a good reference on Kalman filters, see Hamilton 1994). We adopt the Kalman filter methodology to differentiate between consumers' updating their quality valuation at each purchase experience and the time-invariant randomness of their latent utility. Consequently, our model has two main equations. Equation (1) defines the utility of consumer i from choosing brand j at time t (U_{ijt}) as a linear function of the consumer's quality valuation of the brand at that point in time (Q_{ijt}) and a latent utility error. That is,

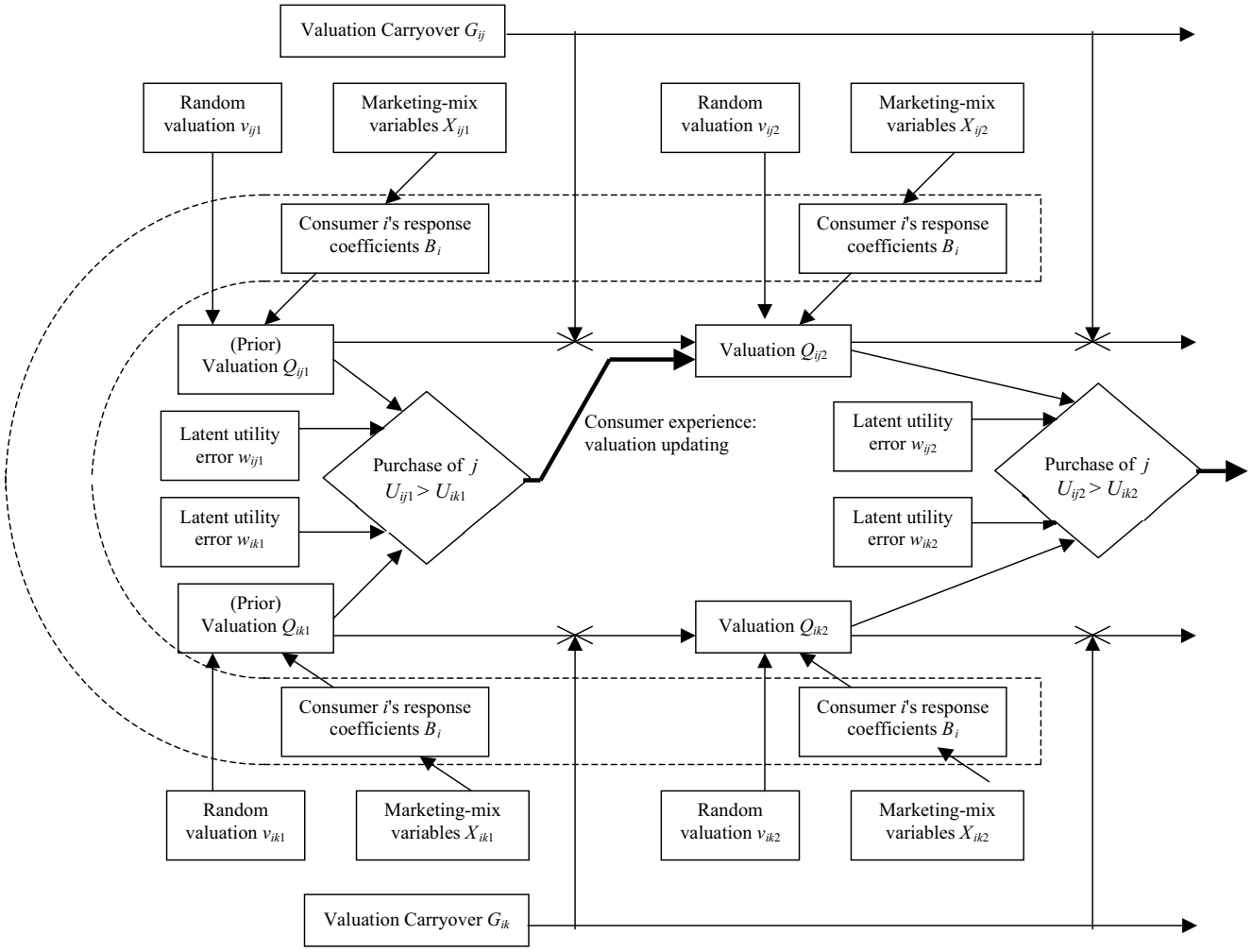
$$U_{ijt} = Q_{ijt} + w_{ijt}. \tag{1}$$

In turn, the quality valuation of a brand encompasses the dynamic aspect of learning and is influenced by the last period's valuation, the marketing-mix variables the consumer is exposed to in the current period, and a random valuation component (v_{ijt}):

$$Q_{ijt} = G_{ij}Q_{ij,\,t-1} + X_{ijt}B_i + v_{ijt}, \tag{2}$$

where G_{ij} measures the impact of the last valuation on the present valuation (brand-specific valuation carryover effect), X_{ijt} denotes a vector of independent

Figure 1 Illustration of Valuation Updating with a Purchase of Brand j



marketing-mix variables, and B_i is the corresponding vector of consumer response parameters.

Note that we distinguish between the seemingly equivalent constructs of valuation and utility because of their distinctly different error structures in the context of consumer learning. Consumers will purchase the brand with the highest utility. Brand utility depends on the quality valuation and a random component (Equation (1)). As is customary in choice models, we call this random component a latent utility error and note that it is not subject to updating. Latent utility error captures missing variables, functional form misspecifications, and idiosyncratic behavior related to purchase-specific factors. In turn, a consumer's quality valuation of a brand, Q_{ijt}, is based on the brand's past valuation, marketing-mix variables, and an unobserved valuation component (Equation (2)). A consumer's quality valuations reflect learning through experience and are subject to Bayesian updating at each purchase experience. Thus, by decoupling the two types of errors in the dynamic choice decision process, we can track the evolution of valuations through updating. Furthermore, by disentangling the components of the overall variance, we can identify the latent utility error by separating it from the random valuation (Hamilton 1994).

The process of valuation learning through updating may exhibit signs of carryover decline due, for example, to forgetfulness over time, brand name confusion, diminished salience, or inconspicuousness of the drug's performance. Hereafter, we use the general term carryover decay to indicate the magnitude of decline in valuation carryover from its perfect-recall value of 1, regardless of the actual source of the decline (thus, we define carryover decay as the value of $1 - G_{ij}$, where G_{ij} is nonnegative).

Note that a G_{ij} coefficient valued at 0 corresponds to a lack of valuation carryover and, therefore, to absence of learning. If the G_{ij} coefficient is negative, the consumer may either have had a treatment experience with brand j that has fallen short of expectations or be prone to switching regardless of the brand's

effectiveness. If the value of this coefficient is positive but less than 1, there is evidence of forgetfulness, or carryover decay. A relatively high value of G_{ij} reflects a high carryover effect and suggests that if a consumer has salient and positive memories about the brand, she is more likely to repeat the purchase in the future. A G_{ij} coefficient of 1 indicates that the consumer remembers her previous valuation perfectly with no carryover decay.[1] Considering the nature of the OTC drugs category, we expect the carryover coefficients G_{ij} to lie mostly between 0 and 1. We note that in this case, carryover decay will have only positive values.

2.2. Relevant Empirical Research

Extant marketing literature is extremely rich in research on consumer brand choice, learning behavior, and dynamic processes. We present a streamlined summary of the pertinent research that is most closely related to our approach.

Fourt and Woodlock (1960), who study the evolution of aggregate sales over time, are among the first to propose a model in which each successive purchase of a brand increases the probability of a next purchase of the same brand. They suggest that such a model is able to serve as a diagnostic tool for market performance (market failure).

Early brand choice models assume that the components of the marketing mix, along with a consumer's purchase history and a random term, determine a consumer's brand choice (a classic example is Guadagni and Little's 1983 study). A consumer's history itself is captured either by actual past purchases, known as *structural state dependence*, or through habit persistence whereby the relative quality of the previous choice is presumed to determine current choice. Recent examples of modeling purchase history, state dependence, and habit persistence can be found in Seetharaman et al. (1999) and Haaijer and Wedel (2001). The built-in dichotomy between the two separate error components in our model and the selective updating that occurs in just one of them distinguish our framework from those above. Thus, our model permits additional insights, though many commonalities in the spirit behind our approach and the models just mentioned can be found.

In a comparative study of several brand choice models, Keane (1997) cautions against omitting either heterogeneity or state dependence and introduces autocorrelation in the latent error term to capture temporal persistence in brand choice more completely. He demonstrates that models accounting for both market dynamics and unobserved heterogeneity produce less biased estimates. Consistent with this view, our model encompasses dynamic purchase behavior through valuation updating and incorporates unobserved consumer heterogeneity.

Previous researchers have advocated the use of Bayesian updating in modeling learning behavior (Zellner 1988). Consumer idiosyncrasies can be accounted for in a Bayesian framework even in the case of only a few longitudinal data points per consumer. Allenby and Lenk (1994) adopt the hierarchical Bayes method to estimate brand choice behavior and impose an autocorrelation structure to capture dynamic choice behavior over time. However, actual learning behavior has been found to differ from the forecasts derived in a Bayesian framework. Studies have shown that people tend to discount the newly acquired information to a higher extent than analytically optimal, thus ascribing a greater weight to their priors (Hoch and Deighton 1989, Philips and Edwards 1966). In our model, we acknowledge the individual-specific learning differences and estimate individual discounting factors in a Bayesian learning framework, thereby allowing flexibility in the process of accommodating new information.

Xie et al. (1997) apply a modified version of Kalman filters to diffusion models. The specific context of their model calls for aggregate estimation of the parameters of interest, whereas we provide individual-specific estimates obtained by hierarchical Bayes analysis in an entirely different marketing context.

Paap and Franses (2000), hereafter called P&F, model brand purchases in a multivariate probit model, incorporating lagged utility and consumer heterogeneity along with marketing variables such as price, display, and feature. The structural resemblance between our framework and P&F's warrants a clearer distinction between the two models.

First, even though there is similarity between our carryover coefficients G_{ij} and the corresponding (Π) matrix in P&F, in our model the carryover occurs in the valuation function, not in the utility function, as in P&F. Consequently, our model allows for a finer specification of the error structure, whereby we separate the lagged latent utility from the current latent utility and random valuation errors. Therefore, our model uses past unobserved information that is embedded in previous choices to adjust the individual parameter estimates in a self-adaptive manner, which endows the model with more flexibility. For example, a purchase-specific factor, such as the temporary unavailability of a favored brand and the attendant information contained in the choice of a less-preferred substitute, will change the dynamics in our model and help fine-tune the individual parameters further.

[1] The value of G_{ij} cannot exceed 1 because it is time-invariant; thus, it will be subject to regression to the mean in the case of high sporadic valuations (i.e., if there were time-dependent carryover coefficients in the model, we would have expected single G_{ijt} values that exceeded 1).

Second, P&F explicitly incorporate lagged effects of marketing-mix variables in the current utility function, whereas we subsume them within the valuation terms and thus subject them to the same forgetting process. Third, our carryover coefficients G_{ij} are individual-specific and thus account for idiosyncrasies in the brand-related carryover effects. Fourth, our model arises solely from dynamic updating of valuation through learning, whereas P&F incorporate base preferences through brand-specific intercepts.

These distinctions should be considered in the context of the product categories the two models analyze: In P&F's case, the product under study is saltine crackers, whereas we study OTC drugs. We believe that in our case, the significance of consumption in the category might induce an enhanced perception of meaningful brand differences, justifying a different model of evaluation and learning. Our model rests on the premise that consumer learning will assume greater relevance with OTC drugs, to the extent that learning will dominate the long-term effects of promotion and render base preferences of lesser consequence. Therefore, the two models are not nested and arise from subtle differences in our understanding of the processes that drive consumer behavior in the two product categories.

2.3. Model Specification and Distributional Assumptions

To identify the model outlined in §2.1, we need to include the no-purchase option (or the outside good) as our base alternative. Thus, we add the no-purchase option $(J+1)$ and define it by the elapsed time parameter, τ_{it}, which indicates interpurchase time and is reset to zero upon a purchase:

$$Q_{i,J+1,t} = \alpha_i \tau_{it} + v_{i,J+1,t}. \tag{3}$$

During the no-purchase state, the consumer could be feeling well with no need for medication, using up the inventory at home, or seeking alternative treatments such as massage, acupuncture, home remedies, or natural supplements, none of which is tractable by scanner data. We refrain from more involved structures because of the already complicated estimation procedure. Hereafter, we refer to α_i as the no-purchase time coefficient and note that, if positive, it indicates a decreasing purchase probability over time. This allows for yet another meaningful interpretation of α_i as an indicator of a consumer's purchase disinclination, which we use alternatively to underscore this particular connotation.

To estimate the full variance-covariance matrices, in the estimation of Equation (2) we use the differenced quality valuations by subtracting the quality valuation of the $(J+1)$ no-purchase option from the quality valuations of the J brands. Similarly, during the estimation, the utility equation (Equation (1)) incorporates the corresponding differenced parameters from the valuation equation. With the no-purchase option serving as the base alternative, if consumer i is faced by negative U_{ijt}'s at time t for $j = 1, \ldots, J$, she chooses the no-purchase option; otherwise, she opts for the brand with the highest positive utility.

Let v_{it} and w_{it} represent vectors obtained by stacking v_{ijt} and w_{ijt}, where $v_{it} \sim N_J(0, V)$, $w_{it} \sim N_J(0, W)$, and V and W are $J \times J$ covariance matrices. The normal distribution assumptions make derivations easier, because the sum of normal random variables is also normally distributed. Although v_{it} and w_{it} may be uncorrelated, for the sake of model flexibility we do not impose any such restrictions during the estimation process and estimate the full probit model with no constraints on the error structure. Thus, we assume $E(v_{it} w_{it}') = \Gamma$, where Γ is a $J \times J$ covariance matrix.[2]

Let B_i be a $k \times 1$ vector that represents the response coefficients and the no-purchase time coefficient. The prior distribution of the valuation function (Q_{it}) given B_i and G_i is assumed to be multivariate normal with dimension J,

$$N_J(G_i M_{i,t-1} + X_{it} B_i,\ G_i C_{i,t-1} G_i' + V), \tag{4}$$

where G_i is a diagonal matrix that contains the G_{ij} coefficients along the diagonal, $M_{i,t-1}$ is a vector of means, and $C_{i,t-1}$ is the variance–covariance matrix of consumer i's valuation at time $t-1$. The purchase dynamics are as follows: A purchase decision (followed by usage) in a given period alters the consumer's prior distribution of the valuation, yielding a posterior distribution. This posterior distribution serves as the prior distribution in the next period. The posterior distribution of the perceived valuation function for consumer i at time t is $N_J(M_{it}, C_{it})$, where

$$M_{it} = (G_i M_{i,t-1} + X_{it} B_i) + \mathrm{Cov}(Q_{it}, U_{it}) \mathrm{Var}(U_{it})^{-1} \cdot \left[U_{it} - (G_i M_{i,t-1} + X_{it} B_i)\right], \tag{5}$$

$$C_{it} = \mathrm{Var}(Q_{it}) - \mathrm{Cov}(Q_{it}, U_{it}) \mathrm{Var}(U_{it})^{-1} \mathrm{Cov}(Q_{it}, U_{it})', \tag{6}$$

where $\mathrm{Cov}(Q_{it}, U_{it}) = G_i C_{i,t-1} G_i' + V + \Gamma$, $\mathrm{Var}(U_{it}) = G_i C_{i,t-1} G_i' + V + W + \Gamma + \Gamma'$, and U_{it} is the utility vector, the components of which are drawn by Gibbs sampling.[3]

[2] Note that even though we do not impose independence between the individual-specific error terms v_{it} and w_{it}, the estimation results indicate that they are independent. We demonstrate and discuss this subsequently.

[3] We obtain the mean and the variance associated with the posterior distribution by standard methods (see, e.g., DeGroot 1970, Mike and Harrison 1989).

In the case of a product trial that is not followed by repeat purchase, the valuation mean and variance evolve as per Equation (4). The valuation mean (the expected value of the brand valuation) decreases over time as a result of carryover decay. In contrast, the valuation variance increases over time, because there is no new signal from another experience. Thus, over time, the consumer begins to forget, the uncertainty (the posterior variance) of the retained valuation increases (even though it is unlikely to reach its prior level), and the posterior mean of the brand valuation regresses to a stationary value.

Now suppose that consumer i subsequently buys the same brand j at time t. There is now new information and valuation updating at time t. Assume that the current experience with the brand leads to positive assessment. The valuation mean increases because of the perceived positive experience (Equation (5)), whereas the valuation variance decreases because of the new information that is brought in by the experience (Equation (6)). That is, at time t, consumer i receives positive reinforcement of her earlier valuation, which she will hold with higher certainty in subsequent periods.

Experience can have a positive or a negative impact on a consumer's mean valuation. For example, if the second term in Equation (5) is negative, then the consumer is adversely affected by the experience. Consequently, the posterior mean will decline as a result of unmet or exceptionally high expectations. Thus, a purchase incidence may increase or decrease the mean of the valuation function for a given brand. However, each experience reduces the consumer's uncertainty (posterior variance) because it reveals extra information about the perceived brand efficacy. This is shown in Equation (6), in which the second term (after the minus sign) is always positive. Therefore, each repeated experience lowers the variance of the valuation function for the chosen brand.

The amount of new information obtained through repetitive usage will decline over subsequent drug applications, and the consumer will hold the developed brand valuation with near certainty, a characteristic of ongoing learning. The nature of this learning is essentially a Bayesian updating process that results in shrinkage around an increasingly stationary mean of the consumer's brand valuation and in decreasing variance of the valuation.

Learning takes place subject to a discounting term. In Equations (5) and (6), the changes in the posterior mean and in the variance are discounted by the term

$$\begin{aligned}&\mathrm{Cov}\left(Q_{it}, U_{it}\right)\mathrm{Var}(U_{it})^{-1}\\ &\quad=\left(G_i C_{i,t-1} G_i' + V + \Gamma\right)\\ &\qquad\cdot\left(G_i C_{i,t-1} G_i' + V + W + \Gamma + \Gamma'\right)^{-1}.\end{aligned}$$

For instructive purposes, we note three special cases of the posterior distribution:

(i) If the latent utility error variance (W) approaches infinity (as V remains the same), no learning takes place. Observe that as the latent utility error variance becomes infinite, the discounting term approaches zero. In this case, the posterior mean equals the first term in Equation (5), and the posterior variance equals the first term in Equation (6). Thus, the distribution of the valuation does not change and no learning takes place.

(ii) If the latent utility error is not specifically accounted for, the model collapses to a restricted version in which choice behavior is captured by Equation (2) only. This follows because $U_{ijt} = Q_{ijt}$ when $w_{ijt} = 0$ in Equation (1), and thus v_{ijt} in Equation (2) captures all the randomness. In this case, there is no updating and the model cannot capture learning. We subsequently estimate and compare this restricted version with the full model.

(iii) Because the carryover decay varies across brands for each consumer, the discounting terms also vary, capturing the heterogeneity during information updating. As carryover decay decreases and the G_{ij} coefficients become closer to 1, the weight of new information during the updating process increases. We test the assumption of near-perfect recall, which is manifested as limited, homogeneous carryover decay, by setting each G_{ij} to 0.95.

Our data set does not contain personal information pertinent to our research objectives. For example, we do not observe the consumer's condition. That is, we do not know whether the buyer is a fitness enthusiast with a strained muscle or a person in rehabilitation after a car accident. Moreover, we do not know the consumer's diet, lifestyle, family history, concurrent ailments, or preexisting conditions. To capture such idiosyncrasies in a framework of valuation updating through learning, we estimate the parameters using hierarchical Bayes analysis in conjunction with Gibbs sampling (Gelfand and Smith 1990). This approach helps us overcome the limited number of observations per consumer and estimate the individual parameters by pooling information from the whole sample. Next, we present the remaining distributional assumptions that are necessary for Gibbs sampling and outline the structure behind the valuation carryover coefficients and the initial valuation parameters.

Let $G_{ij} = G_j^{\text{Fixed}} + G_{ij}^{\text{Random}}$ and $Q_{ij0} = Q_{j0}^{\text{Fixed}} + Q_{ij0}^{\text{Random}}$ for $j = 1, \ldots, J$. The $J \times J$ matrices G_i, G^{Fixed}, G_i^{Random} are obtained by placing G_{ij}, G_j^{Fixed}, G_{ij}^{Random} on the diagonals of $J \times J$ zero matrices, and the $J \times 1$ vectors Q_{i0}, Q_0^{Fixed}, Q_{i0}^{Random} are obtained by stacking Q_{ij0}, Q_{j0}^{Fixed}, Q_{ij0}^{Random} for each j where $j = 1, \ldots, J$. The prior of the diagonal of G^{Fixed} is distributed $N_J(S, Z)$, the prior of the diagonal of G_i^{Random} is distributed

$N_J(0, N)$, the Q_0^{Fixed} prior is distributed $N_J(P, R)$, and the Q_{i0}^{Random} prior is distributed $N_J(0, F)$.

Furthermore, let $B_i = B^{\text{Fixed}} + B_i^{\text{Random}}$, where B_i is a $k \times 1$ vector. Then, B^{Fixed} captures the fixed effects of the responses. The prior of B^{Fixed} is distributed $N_k(H, L)$, and the B_i^{Random} prior is distributed $N_k(0, K)$.

We assume a hierarchical structure on the covariance matrices of random effects. We build the hierarchy on the joint distribution of the quality valuation and latent utility errors. Let Ω be the $2J \times 2J$ joint variance–covariance matrix given by

$$\begin{bmatrix} V & V+\Gamma \\ V+\Gamma' & V+W+\Gamma+\Gamma' \end{bmatrix}.$$

We assume the prior of Ω is IW_{2J} (5I, 5). Furthermore, we let the prior of N be IW_J (3I, 3), the prior of K be IW_k (3I, 3), and the prior of F be IW_J (3I, 3).[4]

Assuming that D denotes the observations in the data, we represent the full posterior distribution of all the parameters by $[D, U, Q, G, B, Q_0, N, F, K, \Omega]$ using notation from Gelfand and Smith (1990) as follows:

$$\begin{aligned}&[D \mid U, Q]\,[U, Q \mid G, B, Q_0, N, F, K, \Omega] \\ &\quad \cdot [Q_0^{\text{Random}} \mid Q_0^{\text{Fixed}}, F]\,[Q_0^{\text{Fixed}}]\,[F]\,[B^{\text{Random}} \mid B^{\text{Fixed}}, K] \\ &\quad \cdot [B^{\text{Fixed}}]\,[K]\,[G^{\text{Random}} \mid G^{\text{Fixed}}, N]\,[G^{\text{Fixed}}]\,[N]\,[\Omega].\end{aligned}$$

Gibbs sampling yields a sample of the joint distribution of the parameters. It proceeds by iteratively sampling from the parameters' posterior conditionals. We use the most recent sampled values during each step while forming the chains. We draw from each of the posteriors respectively and use a mini-Gibbs step for generating draws from the utilities. We present the estimation details in the Appendix. As part of the initial conditions, we set the initial chain values to zero, except for the variances, which we set to unity.

Identification of the intercepts is a known problem with discrete choice models. A typical approach is to normalize the intercept of one choice alternative to zero for identification purposes. Note that because of the dynamic nature of our model, it does not contain intercepts except on the first purchase occasion we observe. In that case, the products of the initial values of Q_{ij0} and their coefficients (G_{ij}) assume the role of brand-specific intercepts for the brand-specific options, that is, Brands 1 through J, and the model is identified by the lack of a lagged term in the $(J+1)$th (no-purchase) option (see Equation (3)). This is similar to the lack of one intercept in a typical discrete choice model. To identify the full variance–covariance matrix Ω, we scale the Gibbs samples so that the last element on its diagonal, $\Omega_{2J,2J}$, equals unity.

We start by model simulation on synthetic data to test the validity of the model and the robustness of the estimation procedure. The fit between the synthetic "true" parameters and the estimated ones is good, attesting to the model validity and the robustness of the estimation (details are available on request). The closeness between the parameters recovered by the model and their original values indicates that the model captures the underlying process well and enables us to apply it on real data.

3. Estimation

3.1. Data

We obtained an A. C. Nielsen data set on OTC purchases of leg-and-back pain relievers. The data span a period of two years (November 1993–October 1995) and were collected in a syndicated scanner panel study administered in various cities in the United States. Our data set comprises 69 regular consumers in this product category; we define "regular" as consumers making at least four purchases during the two years of the study. This restriction screens out consumers with insufficient opportunities to learn, which is essential for our analysis.[5]

The maximum interpurchase time in the data sample is 30 weeks. At the time of the data collection, the studied product category consisted of 4 major brands and 11 minor brands.[6] We aggregated the minor brands in a separate, "Other Brands" position (hereafter, Brand 5). The data set lacks price information on any of its constituent brands. The total number of observations including the no-purchase occasions is 3,519. The average interpurchase time is 7.3 weeks. Descriptive statistics of the data are presented in Table 1.

On the basis of our data, we observe that Brand 2 is both the leading and the most expensive brand. The

[4] To test the sensitivity of results to the priors of the variance–covariance matrices, we also use alternative priors with IW(5I, 5). The results are insensitive to the choice of priors.

[5] We recognize that the limited nature of our data set does not allow for definitive national market inferences. However, note that because all our market information (including the brands' market shares) is based solely on the purchases consumers made in our data sample, the brand characteristics inferred from their valuation updating and the interpretation of the findings within the market they constitute provide consistency and internal validity. We believe that confining ourselves to inferences derived from and related to the available data does not detract from the explanatory power, the reliability, or the validity of the results as long as they are construed within the sample as a market. Therefore, we regard our insights as sufficiently meaningful, even though their out-of-sample representativeness remains questionable. Our goal is to illustrate the model capabilities, and the data at hand serve this purpose well.

[6] For confidentiality reasons, we cannot disclose the actual brand identities.

Table 1 Descriptive Statistics

	Purchase Frequencies	Market Share (%)	Price (Cents per Tablet)	
			Mean	Std. Dev.
Brand 1	103	21.37	19.11	2.22
Brand 2	125	25.93	24.89	1.36
Brand 3	60	12.45	22.80	0.90
Brand 4	61	12.66	22.01	1.24
Brand 5 (others)	133	27.59		
No-purchase	3,037			
Total	3,519	100		

second-largest market share brand is Brand 1, which is also the cheapest. Brands 3 and 4 are close in both market share and unit price. The only information on brand promotions contained in the data is values of redeemed coupons, but the rates of coupon redemption are fairly low for all four major brands. Because of the limited extent of coupon redemption in our data set, we choose to ignore coupons as a marketing-mix variable and focus on the actual (net) prices paid instead. We do not have information on in-store displays or brand advertising from the time the data were collected.

Consistent with the model outlined and discussed in §2.3 and without loss of generality, we use a six-alternative model, in which Brands 1–4 correspond to the four major OTC drug brands, Brand 5 is a composite brand representing Other Brands, and the sixth (no-purchase) option is the base alternative. Because we deal with differenced quality valuations, this structure yields five equations in total. Specifically, we construct the valuation functions as follows:

$$Q_{ijt} = G_{ij}Q_{ij,\,t-1} + b_i \ln(Price_{jt}) - \alpha_i \tau_{it} + v_{ijt} \quad \text{for } j = 1, \ldots, 4, \quad (7)$$

$$Q_{i5t} = G_{i5}Q_{i5,\,t-1} + \gamma_i \ln(t) - \alpha_i \tau_{it} + v_{i5t}, \quad (8)$$

where $j = 1, \ldots, 5$ denotes the brands, t is the time in weeks, and τ_{it} represents the elapsed time in weeks since the last purchase of consumer i at time t. The random terms for the no-purchase alternative are set to zero and differenced out.

The composite nature of Brand 5 prevents us from using specific marketing-mix variables. Therefore, for purchases made in the composite Other Brands category (Brand 5), we let a time-trend variable capture any market changes pertinent to its constituent brands. Note that in the specification of our model, we adopt logarithmic transformations of the price and time-trend variables to capture diminishing marginal effects.

3.2. Goodness of Fit and Validity

To test the fit of the full two-equation Kalman filter model, we compare it with three restricted versions (Table 2). We estimate a nested version with no latent utility errors (i.e., we assume that the w_{ijt}'s are equal to 0), whereby the two distinct model equations collapse into Equation (2) and $U_{ijt} = Q_{ijt}$. In addition, we test two modifications of this nested model: one with imposed near-perfect valuation carryover and the other with unconstrained valuation carryover.

In the most restricted version (Model 1), the latent utility error and the consumer's random valuation component are collapsed into one random term, and the valuation carryover coefficients are fixed at 0.95 (a value near 1). This model yields a minimum sum of squared errors of 765.87.[7] If we introduce the latent utility error separately from the consumer's random valuation component while keeping the valuation carryover coefficients fixed at 0.95 (Model 2), the minimum sum of squared errors decreases to 745.21. In Model 3, in which we remove the constraint on the valuation carryover (i.e., we introduce free estimation of the valuation carryover coefficients) but ascribe all variability to the random valuation component, the sum of squared errors improves to 463.11. The full model, in which the valuation carryover coefficients are unconstrained and the latent utility error is separated from the random valuation component (Model 4), yields a sum of squared errors of 398.32, indicating a better fit for the model we propose compared with the more restricted ones.

We also note that among the four models estimated previously, the minimum sum of squared errors decreases the most when we remove the valuation carryover constraint, thus allowing greater carryover decay into the model. Separating the latent utility error from the random valuation component further improves the model fit.

As an additional check on model validity and to ensure the robustness of the parameters to the data, we estimate the model on ten randomly selected subsets of the original sample of 69 consumers, with replacement. The posterior means of the model

[7] We compute the sum of squared errors in the following way: During the Gibbs iterations, we find the one-step-ahead predicted choice probabilities for each of consumer i's alternatives at period t. This yields a 6×1 matrix of probabilities for consumer i at period t. Then, we subtract the observed choice at time t and take the squares of all terms in the differenced matrix. Note that at each period, only the observed choice has a probability of one and the rest are zero. To compute the sum of squared errors for consumer i, we sum over all t for $t = 1, \ldots, T_i$. Then, we add up these squared errors over all consumers for $i = 1, \ldots, I$. This produces the sum of squared errors at a given iteration. We average the sums of squared errors and use the averaged sum as an indicator of the model's goodness of fit.

Table 2 Model Goodness of Fit vs. Nested Models

	$w_{ijt} = 0$ (No Latent Utility Errors, Error Terms Collapsed Into One Random Term)	$w_{ijt} \neq 0$ (Latent Utility Errors Estimated Separately from the Random Valuations)
All G_{ij} fixed at 0.95 (proxy for lack of carryover decay)	Model 1 SSE = 765.87	Model 2 SSE = 745.21
No constraint on the G_{ij} (carryover decay enabled)	Model 3 SSE = 463.11	Model 4 (full model) **SSE = 398.32**

Note. SSE = Sum of squared error.

parameters are robust with respect to the sample composition. (The results of these alternative sample estimations are available on request.)

Even though we do not impose independence between the individual-specific error terms v_{it} and w_{it}, the estimation results strongly indicate that they are practically uncorrelated and, because of the normality assumption, independent (Table 3). This finding is significant because of the specific consumer learning context we apply the model to and the separation of the error terms that underpins it. In §2.1, we discuss the distinctions between the random components in Equations (1) and (2) and lay out arguments for their essentially disparate nature and role in the learning process. We now provide empirical support for the posited differential behavior of the random valuation and latent utility components with regard to the updating. Next, we proceed with the posterior parameter estimates and their interpretations.

3.3. Results

The posterior parameters are presented in Table 4.[8] Although the mean value of the price coefficients is negative (−0.185, Table 4), it is not significant, and there is evidence of heterogeneity in the price sensitivity of the sample. Approximately 21% of consumers in our sample have positive price coefficients, 29% seem indifferent to price, and 50% exhibit various degrees of price sensitivity. The finding that half of the consumers exhibit a lack of price sensitivity (as evidenced by close to 0 or positive price coefficients) can be attributed to three factors: the price-as-a-signal-of-quality effect, a degree of loyalty to a drug that consumers perceive to be efficacious, and the infrequent and relatively small expenditure consumers incur in a purchase context that they deem of high personal importance.[9]

Although the magnitude of demonstrated lack of price sensitivity in 50% of the sample is unexpected, the specific product category we study must be considered. The personal importance of the purchase and the difficulty of product feature comparisons might motivate consumers to view the higher price as a sign of better quality or higher drug potency. In this case, a higher price might be perceived to be indicative of speedier recovery and preferred as a trade-off for reduced future costs (both physical and monetary). Alternatively, consumers who develop preference for or loyalty to a brand might be less susceptible to price-induced switching. This insight has found support in other studies (e.g., Krishnamurthi and Raj 1991 find that brand loyalty is negatively related to price sensitivity).

In addition, the sporadic need for a drug from the pain-reliever category, combined with the relatively low out-of-pocket expense a single purchase incurs, might induce consumers to disregard price in favor of perceived benefits.[10] To test this supposition, we estimate the correlation between the no-purchase time coefficients α_i and the price coefficients b_i and obtain a value of 0.4. As we mentioned, α_i can be interpreted as a consumer's purchase disinclination, and the estimated correlation suggests a positive association between consumers' frequency of purchases and their price sensitivity. This supports our conjecture. In summary, we hypothesize that the presence of a large number of consumers who are brand loyal, infrequent category users, or driven by quality inferences based on price helps explain the lack of price sensitivity in half of our sample. This finding underscores the need

[8] Although the chains seem to converge even after 25,000 iterations, we let the computations run for 500,000 iterations. For identification purposes, we scale the price and time coefficients by $\sqrt{\Omega_{10,10}}$ and the variance–covariance matrix elements by $\Omega_{10,10}$. We use the last 100,000 observations for our analysis.

[9] Note that our inferences about consumers' price sensitivity apply to price sensitivity across brands only, and do not account for consumers' experience with the brand. Thus, it will be incorrect to interpret them as indicating that a temporary price promotion of a brand may reduce the choice probability of a consumer who, through experience, has established a high valuation of the brand. Our model does not contain brand-specific price sensitivity; neither does it control for brand experience in the price effects. We find evidence that a large fraction of consumers do not perceive a higher price (relative to competitive brands) as a disincentive for purchase, *ceteris paribus*. Therefore, we propose that price promotions should be reconsidered if their primary purpose is to induce trial or switching in this category. This is an interesting finding with immediate profit implications that requires further research.

[10] A typical 24-tablet package of the studied nonprescription drugs costs between $4.59 and $5.97 at the mean unit prices shown in Table 1.

Table 3 Posterior Joint Covariance Matrix

	Brand 1	Brand 2	Brand 3	Brand 4	Brand 5
Brand 1	0.04	0.04	0.04	0.02	0.01
Brand 2	0.03	0.05	0.02	0.02	0.01
Brand 3	0.03	0.02	0.03	0.03	0.00
Brand 4	0.02	0.01	0.02	0.02	0.02
Brand 5	0.00	0.00	−0.02	0.01	0.00

for brand-specific price simulations to evaluate the overall effect of price discounts.

In a recent empirical study on prescription drugs, Gonul et al. (2001) find evidence that physicians demonstrate a lack of price sensitivity in their choice of prescribed medication, irrespective of patient's insurance coverage. The directional association in the two studies leads us to believe that with drugs, individual-specific characteristics pertinent to the product category may override the customarily expected decrease in purchase probability in case of a higher price. This finding seems to hold for a considerable fraction of the market regardless of the identity of the decision maker and the drugs' prescription status. We stress that as well as being category-specific, this is also likely a segment-specific phenomenon that warrants close attention by brand managers.

The time-trend coefficients for Brand 5, γ_i, are predominantly positive, indicating increasing valuation of this brand over time. However, in the aggregate, $\gamma_\cdot$ is not statistically significant (Table 4). This is to be expected given the nature of the composite brand, because disparate marketing effects, which are captured by the time trend, may be canceling each other in the aggregate. The purchase disinclination (no-purchase) coefficients α_i are overwhelmingly negative, which suggests that the purchase probability increases over time for all consumers in the study. This finding is not surprising considering that our sample consists of regular consumers in the leg-and-back OTC drugs category (recall that regular consumers make at least four purchases in the category during the observation period).

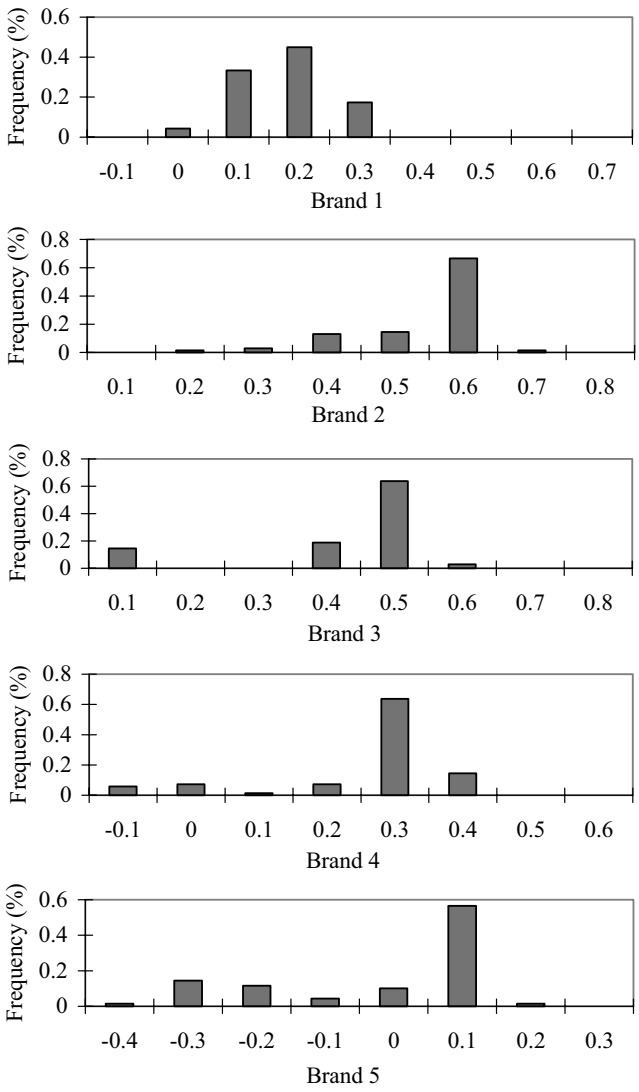
We find that all valuation carryover coefficients (G_{ij}) are less than 1 (Figure 2). Moreover, all their mean values are below 0.5, though Brand 2 has a pronounced mode at 0.6. Thus, we can conclude that a considerable decline in valuation carryover occurs for all consumers and brands. Furthermore, the limited presence of negative valuation carryover coefficients (observed mostly for Brand 5, the composite brand) generally rules out brand dissatisfaction or experimentation as primary motives for switching because of the more systematic, negative impact these factors would have had on the valuation carryover coefficients, which is absent in our case.

The predominance of positive valuation carryover coefficients indicates that most consumers in our sample display a lack of overt switching tendencies. Therefore, the observed cases of brand switching are most likely induced by either carryover decay or marketing-mix effects. This finding, in conjunction with the uncovered low price sensitivity in the sample, leads us to propose that promotional tools

Table 4 Posterior Parameter Estimates

Coefficients	Mean	Std. Dev.
Price ($b_\cdot$)	−0.185	0.117
Brand 5 (others) time trend ($\gamma_\cdot$)	0.312	0.288
No-purchase time coefficient ($\alpha_\cdot$)	−0.208	0.084
Brand 1 valuation carryover (G_1)	0.127	0.050
Brand 2 valuation carryover (G_2)	0.490	0.098
Brand 3 valuation carryover (G_3)	0.367	0.086
Brand 4 valuation carryover (G_4)	0.217	0.076
Brand 5 valuation carryover (G_5)	−0.073	0.078

Figure 2 Valuation Carryover Coefficients



such as prominent displays and meaningful advertising can be of greater significance in this category than brand managers realize. However, more research is needed to substantiate this conjecture.

The brands with pronounced low memorability (significant carryover decay) are Brands 1, 4, and 5. Our data indicate that of these, only Brand 1 has a large market share (Brand 5, the composite brand, is not suitable for specific market share interpretations). On the basis of the limited amount of coupon redemption in our sample, we observe that Brands 1 and 4 engage in price promotions. Although it is difficult to unequivocally establish causality in this case, we can speculate that these brands' manufacturers are compensating for the low memorability of their brands (which leads to lower-than-desired market shares) by frequent price promotions, which are presumably devised to boost sales. However, these two brands may need more radical revisions of their marketing strategies as discussed next.

4. Discussion

4.1. Valuation Updating

To analyze the impact of learning on each brand, we calculate the magnitude of the discounting terms

$$\begin{aligned}&\mathrm{Cov}\left(Q_{it}, U_{it}\right)\mathrm{Var}(U_{it})^{-1}\\ &\quad=\left(G_i C_{i,t-1} G_i' + V + \Gamma\right)\\ &\qquad\cdot\left(G_i C_{i,t-1} G_i' + V + W + \Gamma + \Gamma'\right)^{-1},\end{aligned}$$

discussed in §2.3. Note that the discounting terms are always between zero and one, and they capture the impact of updating on consumers' valuations at time t. If these terms are close to one, there may be significant jumps in the valuation. If a single discounting term is close to zero, then the specific purchase experience has had a negligible impact on the overall consumer valuation of the brand. Because all valuation carryover estimates are less than one in absolute value, the discounting terms will converge to steady-state values.

Judging by the substantially lower-than-one discounting terms, we can infer that there is a considerable amount of new information discounting, which underscores the leverage of prior perceptions. We observe brand-specific variation in the discounting terms, which attests to the variability in valuation updating. Brands 2 and 3 enjoy valuation updating the most, followed by Brands 4, 1, and 5, in that order. Therefore, upon a purchase, consumers using Brands 2 and 3 will change their valuations the most. These two brands also exhibit the highest valuation carryover. Thus, our results converge on the notion that Brands 2 and 3 benefit from superior perceived or actual quality, precise positioning, or effective advertising that reminds consumers of their positive experiences and reinforces the obtained positive impressions.

Brands 1, 4, and 5 also undergo valuation updating (all discounting terms exceed 0.5), indicating changes in consumer valuations upon purchase. However, their considerably lower valuation carryover coefficients suggest that despite the presence of updating, these brands either fail to establish a high level of perceived quality or are less effective in sustaining the established valuation level over time. This may point to weak positioning or inefficient communications.

A closer examination of consumers' discounting coefficients averaged across brands reveals that they fall in the range between 0.53 and 0.59, with a mean of 0.55. The results indicate that the consumers in our sample exhibit similarity in the extent to which they discount the new information gained from recurrent experiences in the product category. Consumers still accommodate new information, indicating ongoing learning, but their prior perceptions remain strong and influential in the face of subsequent experiences. This finding is suggestive of confirmatory and assimilation biases, which are well documented in the consumer behavior literature (for a review, see Hoch and Deighton 1989).

4.2. Limitations and Further Research

Because of the relatively limited data set used to estimate the model, the representativeness of our results might raise some concerns. Despite the demonstrated validity and robustness of the model, the brand-specific discussions presented in §§3.1 and 4.1, though consistent with our findings, should be viewed mostly as an illustration of the kind of inference and comparative diagnostics allowed by the model. Richer data sets are needed to identify more precisely brand inefficiencies that impede consumer learning of brand quality.

Advertising may be important for brand recall in that it exerts a reminding and reinforcing effect. However, we do not have information on contemporaneous advertising campaigns that could have affected the brand valuation carryover in our sample. We expect that even with advertising accounted for, the effect of personal experience in a category in which consumer learning is paramount will dominate in the process of valuation updating, rendering advertising relatively ineffective. In other words, we propose that the role of advertising is likely to be primarily informative and suggestive of application specifics and will exert formative impact that is effective mainly before trial. This issue warrants further exploration, provided relevant data are available.

The findings presented here pertain to a single product category, which limits the generalizability of our results. However, this paper introduces a new

methodology to the marketing brand choice literature and does so in a specific context in which learning through experience is considered highly relevant. Thus, we consider our application mostly instructive and our findings illustrative of the model capabilities. The analytic potential inherent in such a model warrants more extensive work, preferably in other categories in which consumer learning is of essence.

5. Substantive Implications of the Model

Our results are suggestive of a lack of or low price sensitivity in the OTC drugs category, which can be explained by brand loyalty, quality inferences based on price comparisons across brands, or infrequent purchases that render the cost outlay of minor significance relative to the perceived product benefits. Subsequent analysis points to the importance of promotional tools such as informative advertising and primary retail-space allocation. However, because of our lack of relevant data, these inferences should be considered tentative and worthy of future exploration.

The developed model enables us to subject the market performance of the studied brands to a greater scrutiny through detailed comparative analysis. In our particular application on OTC drugs, we uncover significant differences in brand-specific memorability. For example, on the basis of our findings, we can speculate that Brands 2 and 3 remain more recognizable and salient after treatment and are more amenable to valuation updating upon purchase, which may explain their better market performance. In contrast, Brands 1 and 4 exhibit pronounced decline in valuation carryover, which can be attributed to marketing deficiencies that range from confusing labeling and usage directions to poor (perceived) quality, uninformative advertising, or tenuous positioning. This attests to the diagnostic capabilities of the model.

Although we acknowledge that the limited nature of our data thwarts more specific problem diagnostics, we presume that most brand managers will not face this problem. Thus, the inferences drawn from our model, given availability of relevant data, can enable companies to develop educational and product awareness programs that are targeted at segments most likely to benefit from extra information about the specificity of a particular OTC brand. For instance, the brand managers of Brands 1 and 4 can focus on designing unambiguous communication campaigns that describe the symptoms the brands are effective for, and point out contraindications and side effects that may inhibit treatment or aggravate consumer's condition. The social effect of the attained consumer awareness will be twofold: improved efficacy of self-treatment and reductions in public health care costs.

The hierarchical Bayes analysis we employ in the model estimation enables us to account for unobserved heterogeneity in learning and in consumer response to marketing-mix variables. It also enables us to identify price-sensitive consumers and estimate individual-specific levels of discounting the new information in favor of prior perceptions. This detailed response information can ultimately lead to better forecasting of market share and brand performance under different price and promotion scenarios. If consumer identification is possible, a more precise targeting of consumers can be accomplished. Therefore, the proposed model allows for tailored marketing strategies, which can attain greater effectiveness than those suggested by less flexible models.

The model can also be used to analyze consumer behavior and learning in different product categories, for example, fast moving consumer goods. However, the appropriateness of such extensions should be contingent on consumers being sufficiently motivated to learn their own preferences through personal experiences. In contrast, product categories characterized by stimulation-driven variety seeking, conformity to fashion trends and status norms, or prevalence of impulse purchasing will not be appropriate for this model.

In sum, the model will be valuable for the identification of brands that consumers regard with ambiguity or confusion, as evidenced by a low rate of repeat purchasing or a large variability in quality valuations. Highlighting the sources of the deficiencies will alert managers to the need and direction of remedial actions, thus improving the brand positioning in the market. The inferences allowed by the model can be used for brand profiling based on consumers' response to marketing-mix variables, brand memorability, and valuation updating. They can be extended to assist in comparative diagnostics between brands and in market performance simulations.

6. Conclusion

Kalman filter methodology allows for an adaptive, comprehensive model that is capable of tracking valuation evolution by separating the effects of dynamic learning from those of marketing-mix efforts. It allows enduring randomness in the choice decision based on purchase-specific factors while extracting information to adjust the individual parameters accordingly. In this paper, we develop a brand choice model that explicitly accounts for product learning through experience. We obtain estimates of individual marketing-mix parameters, while controlling for idiosyncratic brand preferences and learning behavior.

Our methodological contributions are related to the application of a Kalman filter procedure in a discrete choice model that is estimated by the hierarchical Bayes method. Our substantive contributions of potential benefit to marketing decision makers are in the development of a diagnostic tool for brand performance in a competitive market setting. Brand managers can employ the model to pinpoint marketing problems in the case of a dissatisfactory market performance of their brand. The insights obtained from the analysis can be further used for sales simulations and forecasting under different price and promotion scenarios.

We apply our model to OTC drug purchases because this product category is highly conducive to learning due to the prominence of associated benefits that consumers perceive. Prescription drugs may present the next area for application of the model. In general, the relevance of the model, as well as the specific information value it provides in a particular product category, should be determined by marketing decision makers on an occasion-by-occasion basis.

Appendix: Generate U, Q

We define additional parameters to help us during the presentation of the estimation procedure. Let the elements of the Ω matrix be given by $J \times J$ matrices; that is, Ω_{11}, Ω_{12}, Ω_{21}, and Ω_{22} such that

$$\Omega = \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{bmatrix}.$$

We sequentially generate U_{i1}, Q_{i1} to U_{iT}, Q_{iT} to account for autocorrelation and updating. We first generate U_{it}. The conditional distribution of U_{it}, given purchase observation and other parameters, follows a truncated multivariate normal distribution with mean $(G_i Q_{i,t-1} + X_{it}B_i) + \Omega_{21}\Omega_{11}^{-1}(Q_{it} - (G_i Q_{i,t-1} + X_{it}B_i))$ and variance $\Omega_{22} - \Omega_{21}\Omega_{11}^{-1}\Omega_{12}$. We draw from the conditional distribution of each of the alternatives. In this case, the conditional distribution of an alternative follows a truncated normal distribution. By consecutively drawing from the truncated normal distribution of each of the alternatives, we can generate a mini-Gibbs step that will converge to a draw from the truncated multivariate normal distribution. Accordingly, U_{it} is sampled so that it gives the highest utility for the observed brand. If there is no-purchase observation, all brand utilities receive negative values. For $t = 1$, we use Q_{i1} and Q_{i0} to generate U_{i1} from its conditional distribution and use the prior Q_{i2} to generate U_{i2}. Then, using U_{i1} and U_{i2}, we generate the posterior of Q_{i1}. In general, we generate the posterior of Q_{it} by using U_{it} and $U_{i,t+1}$. The posterior of Q_{it} is

$$\begin{aligned}[Q_{it} \mid Q_{i,t-1}, Q_{i,t+1}, U_{it}, U_{i,t+1}] &\propto [Q_{it} \mid Q_{i,t-1}, U_{it}]\,[Q_{i,t+1} \mid Q_{it}, U_{i,t+1}] \\ &\propto \exp\{-\tfrac{1}{2}(Q_{it} - M_{it})' C_i^{-1} (Q_{it} - M_{it})\}.\end{aligned}$$

There are four different posteriors depending on whether there is updating during periods t and $t+1$. Updating occurs whenever a brand is purchased (or utility is observed), whereas there is no updating if no brand is purchased (or no utility is observed). The respective means and variances for the posteriors are as follows:

(i) Updating at $t+1$ only:

$$\begin{aligned} C_i &= \left(\Omega_{11}^{-1} + G_i\left(I - \Omega_{12}\Omega_{22}^{-1}\right)'\left(\Omega_{11}^{-1} - \Omega_{12}\Omega_{22}^{-1}\Omega_{21}\right)^{-1} \cdot \left(I - \Omega_{12}\Omega_{22}^{-1}\right)G_i'\right)^{-1} \\ M_{it} &= C_i\left(\Omega_{11}^{-1}\left(G_i Q_{i,t-1} + X_{it}B_i\right) + G_i\left(I - \Omega_{12}\Omega_{22}^{-1}\right)' \cdot \left(\Omega_{11} - \Omega_{12}\Omega_{22}^{-1}\Omega_{21}\right)^{-1} \cdot \left(Q_{i,t+1} - \left(I - \Omega_{12}\Omega_{22}^{-1}\right)X_{i,t+1}B_i - \Omega_{12}\Omega_{22}^{-1}U_{i,t+1}\right)\right). \end{aligned}$$

(ii) Updating at t only:

$$\begin{aligned} C_i &= \left(\left(\Omega_{11} - \Omega_{12}\Omega_{22}^{-1}\Omega_{21}\right)^{-1} + G_i\Omega_{11}^{-1}G_i'\right)^{-1} \\ M_{it} &= C_i\left(\left(\Omega_{11} - \Omega_{12}\Omega_{22}^{-1}\Omega_{21}\right)^{-1}\left(\left(I - \Omega_{12}\Omega_{22}^{-1}\right)\left(G_i Q_{i,t-1} + X_{it}B_i\right) + \Omega_{12}\Omega_{22}^{-1}U_{it}\right) + G_i\Omega_{11}^{-1}\left(Q_{i,t+1} - X_{i,t+1}B_i\right)\right). \end{aligned}$$

(iii) Updating at t and $t+1$:

$$\begin{aligned} C_i &= \left(\left(\Omega_{11} - \Omega_{12}\Omega_{22}^{-1}\Omega_{21}\right)^{-1} + G_i\left(I - \Omega_{12}\Omega_{22}^{-1}\right)' \cdot \left(\Omega_{11}^{-1} - \Omega_{12}\Omega_{22}^{-1}\Omega_{21}\right)^{-1}\left(I - \Omega_{12}\Omega_{22}^{-1}\right)G_i'\right)^{-1} \\ M_{it} &= C_i\left(\left(\Omega_{11} - \Omega_{12}\Omega_{22}^{-1}\Omega_{21}\right)^{-1}\left(\left(I - \Omega_{12}\Omega_{22}^{-1}\right)\left(G_i Q_{i,t-1} + X_{it}B_i\right) + \Omega_{12}\Omega_{22}^{-1}U_{it}\right) + G_i\left(I - \Omega_{12}\Omega_{22}^{-1}\right)'\left(\Omega_{11} - \Omega_{12}\Omega_{22}^{-1}\Omega_{21}\right)^{-1} \cdot \left(Q_{i,t+1} - \left(I - \Omega_{12}\Omega_{22}^{-1}\right)X_{i,t+1}B_i - \Omega_{12}\Omega_{22}^{-1}U_{i,t+1}\right)\right). \end{aligned}$$

(iv) No updating:

$$\begin{aligned} C_i &= \left(\Omega_{11}^{-1} + G_i\Omega_{11}^{-1}G_i'\right)^{-1} \\ M_{it} &= C_i\left(\Omega_{11}^{-1}\left(G_i Q_{i,t-1} + X_{it}B_i\right) + G_i\Omega_{11}^{-1}\left(Q_{i,t+1} - X_{i,t+1}B_i\right)\right). \end{aligned}$$

At T_i, because there is no $T_i + 1$, the second terms in the equations given under steps (i) to (iv) drop out, and the posteriors depend on $[Q_{iT_i} \mid Q_{i,T_i-1}, U_{iT_i}]$ only.

The next steps involve generating Q_{i0}^{Random}, Q_0^{Fixed}, F, B_i^{Random}, B^{Fixed}, K, G_i^{Random}, G^{Fixed}, N, and Ω, respectively.

References

Allenby, G. M., P. J. Lenk. 1994. Modeling household purchase behavior with logistic normal regression. *J. Amer. Statist. Assoc.* **89**(428) 1218–1231.

DeGroot, M. H. 1970. *Optimal Statistical Decisions*. McGraw-Hill, New York.

Fourt, L., J. Woodlock. 1960. Early prediction of market success for new grocery products. *J. Marketing* **25**(October) 31–38.

Gelfand, A. E., A. F. M. Smith. 1990. Sampling-based approaches to calculating marginal densities. *J. Amer. Statist. Assoc.* **85** 398–409.

Gönül, F., F. Carter, E. Petrova, K. Srinivasan. 2001. Promotion of prescription drugs and its impact on physicians' choice behavior. *J. Marketing* **65**(July) 79–90.

Guadagni, P., J. D. C. Little. 1983. A logit model of brand choice calibrated on scanner panel data. *Marketing Sci.* **2** 203–238.

Haaijer, R., M. Wedel. 2001. Habit persistence in time series models of discrete choice. *Marketing Lett.* **12**(1) 25–35.

Hamilton, J. 1994. *Time Series Analysis*. Princeton University Press, Princeton, NJ.

Hoch, S. J., J. Deighton. 1989. Managing what consumers learn from experience. *J. Marketing* **53** 1–20.

Keane, M. P. 1997. Modeling heterogeneity and state dependence in consumer choice behavior. *J. Bus. Econom. Statist.* **15**(3) 310–327.

Krishnamurthi, L., S. P. Raj. 1991. An empirical analysis of the relationship between brand loyalty and consumer price elasticity. *Marketing Sci.* **10**(2) 172–183.

Mike, W., J. Harrison. 1989. *Bayesian Forecasting and Dynamic Models*. Springer-Verlag, New York.

Paap, R., P. H. Franses. 2000. A dynamic multinomial probit model for brand choice with different long-run and short-run effects of marketing mix variables. *J. Appl. Econom.* **15**(6) 717–744.

Philips, L. D., W. Edwards. 1966. Conservatism in a simple probability inference task. *J. Experiment. Psych.* **72** 346–354.

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

Xie, J., X. Michael Song, M. Sirbu, Q. Wang. 1997. Kalman filter estimation for new product diffusion models. *J. Marketing Res.* **34**(August) 378–393.

Zellner, A. 1988. Bayesian analysis in econometrics. *J. Econom.* **37** 27–50.