

Disentangling Preferences and Learning in Brand Choice Models

Sangwoo Shin

Krannert School of Management, Purdue University, West Lafayette, Indiana 47907, shin58@purdue.edu

Sanjog Misra, Dan Horsky

Simon School of Business, University of Rochester, Rochester, New York 14627
{misra@simon.rochester.edu, dan.horsky@simon.rochester.edu}

In recent years there has been a growing stream of literature in marketing and economics that models consumers as Bayesian learners. Such learning behavior is often embedded within a discrete choice framework that is then calibrated on scanner panel data. At the same time, it is now accepted wisdom that disentangling preference heterogeneity and state dependence is critical in any attempt to understand either construct. We posit that this confounding between state dependence and heterogeneity often carries through to Bayesian learning models. That is, the failure to adequately account for preference heterogeneity may result in over- or underestimation of the learning process because this heterogeneity is also reflected in the initial conditions. Using a unique data set that contains stated preferences (survey) and actual purchase data (scanner panel) for the same group of consumers, we attempt to untangle the effects of preference heterogeneity and state dependence, where the latter arises from Bayesian learning. Our results are striking and suggest that measured brand beliefs can predict choices quite well and, moreover, that in the absence of such measured preference information, the Bayesian learning behavior for consumer packaged goods is vastly overstated. The inclusion of preference information significantly reduces evidence for aggregate-level learning and substantially changes the nature of individual-level learning. Using individual-level outcomes, we illustrate why the lack of preference information leads to faulty inferences.

Key words: Bayesian learning; brand choice; preferences; state dependence; Markov chain Monte Carlo

History: Received: November 14, 2007; accepted: January 4, 2011; Eric Bradlow served as the editor-in-chief and Jean-Pierre Dubé served as associate editor for this article.

1. Introduction

Consumers choose between brands based on their individual preferences, past experiences, and the brands' marketing mix elements. The marketing and economics literatures are replete with examples of studies that investigate the relative impact each of these factors has on the brand choice decision. Although this literature is quite heterogeneous in its findings, what has emerged as a consensus is that the separate identification of these effects is nontrivial. It is now well documented (see, e.g., Heckman 1991) that the effect individual preferences (heterogeneity) and past experiences (state dependence) have on brand choices can be confounded. That is, we know that a failure to adequately account for heterogeneity in preferences may lead to a bias in the effect of state dependence. This in turn may also lead to a bias in the estimates pertaining to marketing mix effects such as price.

In recent years there has been a growing stream of literature in marketing and economics that aims to structurally model the beliefs that consumers have about each brand and the role such beliefs play in

choice decisions. In particular, this literature treats consumers as Bayesian learners and allows them to update their beliefs via signals obtained at each purchase occasion. For implementation purposes, the Bayesian learning framework is embedded within a discrete choice setting that is then calibrated on consumer choice data (see, e.g., Erdem and Keane 1996, Akerberg 2003, Mehta et al. 2003). This approach specifies a structural model of the brand choice process and allows researchers to delve into the underpinnings of brand choice behavior.

Said differently, the Bayesian learning model can be thought of as a framework that allows for higher-order state dependence, albeit in a fairly parsimonious manner. As such, the discrete choice model with Bayesian learning is data intensive and makes disentangling preference heterogeneity and state dependence even more difficult. The key problem, simply stated, is that consumer learning is not fully identified from revealed choice data. This identification problem has been recognized (Erdem and Keane 1996), and the typical solution is to assume a common prior

across consumers or to somehow use past data to calibrate priors. The crux of the issue is that without data on initial conditions, there is no way of identifying the rate and amount of learning for a given consumer. Note that this problem persists even if data were available from the first choice onward because there remains an informational insufficiency, because prior beliefs at the initial point are still unavailable and remain heterogeneous across consumers. Ignoring this missing information and the implicit heterogeneity induces biases in our understanding of consumer learning and will almost surely misrepresent other constructs of heterogeneity. We note that the problem cannot simply be resolved with long panels because such data are not informative about the consumers' initial priors.

What additional information, then, can researchers collect that would help resolve these issues? The answer is simply to augment revealed preference data with information on consumer preferences. Ideally, if a consumer's true preferences for the relevant brands were available at each purchase occasion, the researcher could pin down the precise underlying behavior that drives observed choices. Such preference information, however, is tedious and expensive to collect. A second-best alternative is to gather data on preferences at some point prior to choices being observed. Such information substantially reduces the researcher's burden about inferring initial priors by substituting data in place of assumptions. As a result, one is better able to resolve the confounding between learning and preference heterogeneity by allowing this augmented preference information to offer a competing explanation for the observed choice sequence.

For instance, if a consumer's preference data reveal a strong idiosyncratic preference for a particular brand of toothpaste (such as Crest) as well as a high degree of familiarity with the brand, the researcher should be able to rule out a learning-based explanation. Sensitivity to marketing mix variables is also better assessed in the presence of such preference information. Continuing the earlier example, perhaps a price discount on the last shopping trip seems to have induced the consumer to switch from her preferred brand (Crest) to a potentially less preferred one (Colgate). Clearly, the availability of preference information now offers insights into the degree of substitutability between the two brands and would directly inform the degree to which the consumer is price sensitive. If data reveal that Colgate and Crest are equally preferred, it would imply that she is less price sensitive than if the consumer strongly preferred Crest to Colgate. Consequently, the price differential required to induce a brand switch from Crest to Colgate should be smaller in the former case than in the latter, and this has direct implications on the estimates of price elasticity. By a

similar argument, the effect of other marketing mix elements would also be more clearly estimated.

In general, there is widespread agreement that stated preferences are based on the true underlying preferences of the consumer. Conjoint studies routinely use multiattribute utility models to construct estimates of consumers' utility functions and use them to predict choices. The previous literature has also paid some attention to the importance of combining revealed preference data with such stated preference data. Early studies such as Ben-Akiva and Morikawa (1990), Hensher and Bradley (1993), and Horsky and Nelson (1992) investigated the behavioral and cognitive process through which stated preference data are generated, and they explained why such data are predictive of actual market behavior. Using cross-sectional health-care plan choice and survey data, Harris and Keane (1999) showed that incorporation of the attitudinal data leads to a substantial improvement in choice model fit and more precise estimates of all choice model parameters. Horsky et al. (2006) reported similar findings in the context of scanner panel data analysis. The above-cited studies are primarily concerned with choice environments in which decision makers act on full information. There has been, however, a growing trend in recent years to model choice processes in which decision makers act with partial information. Manski (2004) recently discussed identification of such decision processes and concluded that choice data alone do not suffice to infer about the underlying behavior. In the spirit of Manski (2004), we attempt to empirically show how misleading inferences about the consumer's learning process can be, particularly when this process is calibrated using only scanner panel data.

In the current study we estimate a logit-based Bayesian learning model in which learning parameters are allowed to be fully heterogeneous via the augmentation of survey information on consumer preferences and familiarities. By comparing it with the Bayesian learning model calibrated on standard scanner panel data alone, we make a number of substantive contributions to the literature. Our findings enhance the current knowledge about the consumer brand choice process. First, we demonstrate how the inclusion of preference and familiarity information substantially alters our understanding of the brand choice process. In particular, the absence of this information significantly overestimates the amount of aggregate-level learning. Correspondingly, the role of preference heterogeneity is much more pronounced in the presence of survey information. Second, our analysis allows us to take a deeper look at the individual-level choice process and to consequently document the effect that preference heterogeneity and learning have on explaining individual-level purchase patterns. Finally, we find that the inclusion of preference

information uncovers statistically and managerially significant biases in parameter estimates, such as price sensitivity, and the degree of parameter heterogeneity. Although our results are an initial foray into the topic and are based on a single data set and one particular model specification, they do offer the marketing scientist new insights into disentangling the impact of preferences and learning in consumer's brand choice.

The rest of this paper is organized as follows: In the next section, we lay out the Bayesian learning process and illustrate how to embed it in a discrete choice framework. In §3 we describe our unique data set, which combines stated preferences (survey) and actual purchase data (scanner panel) for the same group of consumers in the toothpaste market. We specify how the survey information on preferences and familiarity of the brands is incorporated into the Bayesian learning model. In particular, our specification of the learning process uses additional parameters that allow the consumer to update fully heterogeneous initial preferences. At the same time, we discuss the identification issues associated with estimation of the Bayesian learning model using scanner panel data. We then describe our estimation methodology and follow this with a discussion of our empirical findings. The parameter estimates of the Bayesian learning model that relate to preference heterogeneity, learning, and marketing mix variables are provided. We follow with discussing managerial and research implications of our study. We conclude with a summary.

2. The Bayesian Learning Model

The seminal work of Erdem and Keane (1996) has generated a stream of papers in marketing and economics that incorporate the Bayesian learning process into a discrete choice framework.¹ These models have been used to model choices in various application areas, from consumer's brand choice decisions (e.g., Erdem and Keane 1996, Akerberg 2003, Mehta et al. 2003) to physicians' prescription decisions (e.g., Crawford and Shum 2005, Narayanan and Manchanda 2009). In keeping with this literature, we will assume that beliefs are updated via a Bayesian learning mechanism with normal priors and signals. In addition, we will also assume that consumers are risk neutral and myopic.

In the rest of this section, we lay out the model specification and main assumptions underlying the Bayesian learning model. Our exposition in the sequel focuses mainly on the Bayesian learning model in the context of its estimation using scanner panel data.

2.1. The Bayesian Quality Learning Process

In the framework described below, consumers learn about brand quality by updating their beliefs over successive purchase occasions. More specifically, consumers receive a quality signal after each purchase, combine the information contained in this signal with their prior beliefs, and construct a posterior belief in accordance with Bayes's rule. In this context, "learning" is conceptualized as having two distinct effects: *bias reduction* and *uncertainty reduction*. The first effect stems from the stochastic convergence of a consumer's quality perception to the true mean quality (bias reduction), whereas the second effect reflects the deterministic convergence of uncertainty to zero (uncertainty reduction). This two-dimensional nature of the Bayesian learning process yields a parsimonious yet flexible learning mechanism.

Let $Q_{ij,t}^s$ denote a signal about brand j 's quality that consumer i receives after purchasing brand j at time t . We assume that quality signals are generated from the following normal distribution:²

$$Q_{ij,t}^s \sim N(Q_{ij}, \sigma_{Q_{ij}}^2), \quad (1)$$

where Q_{ij} is consumer i 's true mean quality assessment (or match value) of brand j , and $\sigma_{Q_{ij}}^2$ is the signal variance of brand j faced by consumer i . Given that $\sigma_{Q_{ij}}^2 > 0$, quality signals only contain partial information about the unknown true mean quality. Again, the quality signal is assumed to be realized only after consumer i purchases and consumes brand j at time t .

Consumer i is assumed to have an initial quality belief about the unknown true mean quality of brand j , as given below:

$$\tilde{Q}_{ij,0} = N(\mu_{Q_{ij,0}}, \sigma_{Q_{ij,0}}^2). \quad (2)$$

In the above equation, $\mu_{Q_{ij,0}}$ and $\sigma_{Q_{ij,0}}^2$ are initial beliefs about the mean and variance of brand j 's quality at time 0, respectively. We note here that in the Bayesian paradigm, prior beliefs at any time t are simply the posterior beliefs at time $t - 1$. In other words, successively combining prior beliefs with the consumption signals allows us to construct the posterior belief at any time $t > 0$. This time-specific posterior belief also follows a normal distribution and is denoted by

$$\tilde{Q}_{ij,t} = N(\mu_{Q_{ij,t}}, \sigma_{Q_{ij,t}}^2). \quad (3)$$

Because quality beliefs at any time $t \geq 0$ are normally distributed, they are completely characterized by their mean and variance parameters. In other

¹ Earlier work in this vein includes Roberts and Urban (1988), Eckstein et al. (1988), and Horsky and Raban (1988).

² We are assuming that non-purchases are uninformative from a quality learning point of view. An alternative specification could include a reinforcement learning component wherein non-purchases also have a role to play, in the spirit of Camerer et al. (2002).

words, the laws of motion for the posterior mean and variance are sufficient to characterize the evolution of a consumer's quality beliefs. If consumer i updates posterior beliefs at time $t - 1$ (or prior belief at time t) through the realization of quality signals in a Bayesian fashion, the posterior mean and variance at time t can be updated in the following recursive manner:

$$\mu_{Q_{ij},t} = \frac{\sigma_{Q_{ij},t}^2}{\sigma_{Q_{ij},t-1}^2} \mu_{Q_{ij},t-1} + y_{ij,t} \frac{\sigma_{Q_{ij},t}^2}{\sigma_{Q_{ij},t}^2} Q_{ij,t}^S \quad \text{and} \quad (4)$$

$$\frac{1}{\sigma_{Q_{ij},t}^2} = \frac{1}{\sigma_{Q_{ij},t-1}^2} + y_{ij,t} \frac{1}{\sigma_{Q_{ij},t}^2}, \quad (5)$$

where $y_{ij,t}$ is an indicator variable such that $y_{ij,t} = 1$ if consumer i purchases brand j at time t and $y_{ij,t} = 0$ otherwise. Successive substitutions of Equations (4) and (5) result in alternative expressions for $\mu_{Q_{ij},t}$ and $\sigma_{Q_{ij},t}^2$ as given by

$$\mu_{Q_{ij},t} = \frac{\sigma_{Q_{ij},t}^2}{\sigma_{Q_{ij},0}^2} \mu_{Q_{ij},0} + \frac{\sigma_{Q_{ij},t}^2}{\sigma_{Q_{ij},0}^2} \sum_{\tau=1}^t y_{ij,\tau} Q_{ij,\tau}^S \quad \text{and} \quad (6)$$

$$\frac{1}{\sigma_{Q_{ij},t}^2} = \frac{1}{\sigma_{Q_{ij},0}^2} + \frac{\sum_{\tau=1}^t y_{ij,\tau}}{\sigma_{Q_{ij},0}^2}. \quad (7)$$

From an estimation standpoint, it is useful to construct an alternative expression of the Bayesian learning process using a change of variables. To do this we define two new variables, $\nu_{Q_{ij},t} = \mu_{Q_{ij},t} - Q_{ij}$ and $\eta_{ij,t}^S = Q_{ij,t}^S - Q_{ij}$. These new variables, $\nu_{Q_{ij},t}$ and $\eta_{ij,t}^S$, are referred to as *perception bias* and *signal noise*, respectively. The former measures how much consumer i 's mean quality perception deviates from the true mean quality, whereas the latter represents a noise component of the quality signal. Using these transformations and combining Equation (7) with (6) lead to the final expression for the mean quality perception, given by

$$\begin{aligned} \mu_{Q_{ij},t} &= Q_{ij} + \nu_{Q_{ij},t} \\ &= Q_{ij} + \frac{(\sigma_{Q_{ij},t}^2 / \sigma_{Q_{ij},0}^2) \nu_{Q_{ij},0} + \sum_{\tau=1}^t y_{ij,\tau} \eta_{ij,\tau}^S}{\sigma_{Q_{ij},t}^2 / \sigma_{Q_{ij},0}^2 + \sum_{\tau=1}^t y_{ij,\tau}}. \end{aligned} \quad (8)$$

The mean quality perception starts with $Q_{ij} + \nu_{Q_{ij},0}$ at the initial period ($t = 0$), evolves over time as suggested in Equation (8) (for $t \geq 1$), and converges to Q_{ij} at steady state ($t = t^*$, where $\nu_{Q_{ij},t^*} = 0$ and $\sigma_{Q_{ij},t^*}^2 / \sigma_{Q_{ij},0}^2 = \infty$). The above equation represents the crux of the Bayesian learning process. It highlights the fact that the mean quality perception $\mu_{Q_{ij},t}$ can be decomposed into two components: a time-invariant Q_{ij} and a time-varying $\nu_{Q_{ij},t}$. The existence of the time-varying component differentiates the Bayesian learning process from the zero-order process. If $\nu_{Q_{ij},0} = 0$ and $\sigma_{Q_{ij},0}^2 = 0$ (therefore, $\nu_{Q_{ij},t} = 0 \forall t$), the Bayesian

learning process collapses to the zero-order process (i.e., $\mu_{Q_{ij},t} = Q_{ij}$). This case describes a consumer who is no longer learning (about brands) because his or her quality perception already converged to the true mean quality, and no uncertainty about his or her quality perception remains.

The unique specification of the time-varying component also differentiates the Bayesian learning process from the alternative approaches of modeling time-varying preferences such as the popular inertia/purchase reinforcement process (i.e., $\mu_{Q_{ij},t} = Q_{ij} + \lambda_i y_{ij,t-1}$). There are two noticeable differences between the inertia and Bayesian learning processes. First, the extent of state dependence is different. The inertia process has only a first-order effect (i.e., only the brand choice lagged by one time period affects the current brand choice decision), whereas the Bayesian learning process is a higher-than-first-order process (which is often referred to as an infinite-order process, in which the entire choice history affects the current brand choice decision). More importantly, the nature of state dependence is different. The effect of inertia is usually modeled as not varying across brands or over time, whereas that of learning is heterogeneous across brands and is diminishing over time.

2.2. Utility Specification

Given the specification of a consumer learning process, we now move to describing the consumers' utility and choice framework. As is typical in discrete choice models, we specify consumer i 's utility from purchasing brand j at time t as the following linear form:

$$U_{ij,t} = \tilde{Q}_{ij,t-1} + X_{ij,t} \beta_i + \varepsilon_{ij,t}^U, \quad (9)$$

where $\tilde{Q}_{ij,t-1}$ is consumer i 's beliefs about brand j 's quality at time t , $X_{ij,t}$ stands for the vector of marketing mix variables of brand j observed by consumer i at time t , β_i is the corresponding vector of response coefficients, and $\varepsilon_{ij,t}^U$ stands for utility components unobserved to researchers. Note that when consumer i makes a purchase decision at time t , the brand quality signal is not yet realized and hence is not considered an observable. In other words, quality beliefs in our notation are lagged by one time period to represent the fact that the quality beliefs updated after purchase occasion at time $t - 1$ are relevant to purchase decision at time t . Since $\tilde{Q}_{ij,t-1}$ is a random variable, consumer i bases decisions on the expected value of utility with respect to quality beliefs. This expected utility can be computed as

$$\begin{aligned} U_{ij,t}^E &= E[U_{ij,t}] \\ &= E[\tilde{Q}_{ij,t-1}] + X_{ij,t} \beta_i + \varepsilon_{ij,t}^U \\ &= \mu_{Q_{ij},t-1} + X_{ij,t} \beta_i + \varepsilon_{ij,t}^U. \end{aligned} \quad (10)$$

Using the alternative parametrization introduced earlier in Equation (8), we can also express the above equation as

$$U_{ij,t}^E = Q_{ij} + \nu_{Q_{ij},t-1} + X_{ij,t}\beta_i + \varepsilon_{ij,t}^U. \quad (11)$$

The assumption that $\varepsilon_{ij,t}^U$ is independent and identically distributed (*iid*) Type I extreme value distributed completes the model specification and leads to the random coefficient multinomial logit model where brand-specific intercepts are composed of true mean qualities (Q_{ij}) and perception biases ($\nu_{Q_{ij},t-1}$). Notice that the model specification presented here is simpler than are other variants of the Bayesian learning model available in the literature. There are several ways to extend the current specification if the researcher so desires. First, the assumption that $\varepsilon_{ij,t}^U$ is multivariate normal distributed leads to the random coefficient multinomial probit model with a full covariance matrix (Narayanan and Manchanda 2009). Second, risk aversion can be incorporated via the Constant Absolute Risk Aversion subutility function (Erdem and Keane 1996, Crawford and Shum 2005, Narayanan and Manchanda 2009). Third, forgetting can be embedded in the learning model by allowing for the possibility of imperfect recall (Mehta et al. 2004). Finally, forward looking can be modeled together with learning (Erdem and Keane 1996, Ackerberg 2003, Crawford and Shum 2005). In this study, we keep the model specification as simple as possible to focus on our main research questions by avoiding methodological and interpretational complications.³

2.3. Identification Issues

In what follows, we informally discuss the identification of the parameters in Bayesian learning models. Identification of the model parameters depends on the specification of the model as well as the data available. The Bayesian learning model presented

in the earlier section is not identifiable in its current form. Some parameters are by design unidentifiable, whereas others are challenging to identify with typical scanner panel data. One therefore needs to impose a number of additional restrictions to achieve identification.

The set of parameters that characterize the Bayesian learning process is $\{Q_{ij}, \sigma_{Q_{ij}}^2, \nu_{Q_{ij},0}, \sigma_{Q_{ij},0}^2\} \forall i$ and j . From Equations (7) and (8) it is obvious that the initial perception variance $\sigma_{Q_{ij},0}^2$ and the quality signal variance $\sigma_{Q_{ij}}^2$ are not separately identified, but only their ratio, $\sigma_{Q_{ij}}^2/\sigma_{Q_{ij},0}^2$, is identifiable. To resolve this, one usually sets $\sigma_{Q_{ij}}^2 = 1 \forall i$ and j . The literature adopts various rules in dealing with these identification issues. Whereas Mehta et al. (2003) impose the same restriction to identify this ratio, Narayanan and Manchanda (2009) use a different strategy in the sense that (i) their initial perception variance is only brand specific (i.e., $\sigma_{Q_{ij},0}^2$) and is drawn from a common prior distribution, and (ii) the quality signal variance is only individual specific (i.e., $\sigma_{Q_i}^2$) and is estimated from data.

Under our specification, the interpretation of estimated $\sigma_{Q_{ij},0}^2$ is deemed relative to $\sigma_{Q_{ij}}^2 = 1$ (e.g., $\hat{\sigma}_{Q_{ij},0}^2 = 1/2$ implies that $\sigma_{Q_{ij},0}^2$ is one-half of $\sigma_{Q_{ij}}^2$). The true mean qualities, Q_{ij} , can be thought of as steady-state individual-level intercept terms. As is typical, not all of these brand-specific Q_{ij} s are identified, and we need to specify a reference brand (e.g., say, $Q_{ij} = 0$).

A remaining question is how one identifies $\nu_{Q_{ij},0}$ and $\sigma_{Q_{ij},0}^2$ (separately from Q_{ij}). These prior quantities represent heterogeneous initial conditions that cannot simply be identified from revealed choice data without strong untestable assumptions. The identification problems arise because typical standard scanner panel data are usually left-truncated (initial conditions) and/or are often relatively short. In addition, typical patterns of choices in frequently purchased product categories make this identification task harder to conduct. First, the market share of brands covaries with changes in brands' marketing mix investments that could mask the systematic evolution of the market share, which is essential to separating the prior bias and uncertainty constructs from the steady-state intercepts. This problem is exacerbated because of the maturity of typical product categories for which scanner panel data are available.⁴ Second, consumers often purchase a brand from a small subset of the brands available in the product category. That is, $y_{ij,t} = 0$ for some j during the entire purchase history

³ The inclusion of forward-looking behavior substantially increases the computational burden because it requires solving a dynamic programming problem in the course of estimation. As a consequence, forward-looking behavior is often introduced at the expense of heterogeneity in the Bayesian learning model. Furthermore, the incorporation of forward-looking behavior is usually undertaken to allow for experimentation, which in the context of a mature product category with experienced buyers seems less important. We also ignore risk aversion in our application. Given that the focus of this study is to discuss the role of survey information augmented to scanner panel data in improving identification of the Bayesian learning model, we conjecture that the findings reported in this study would be robust to the change of model specification. Finally, our results are based on a multinomial logit (MNL) specification. A more general framework, such as the multinomial probit, will be able to capture correlations across brands at each purchase occasion and, consequently, might offer richer and cleaner insights into learning.

⁴ This problem is mitigated by relying on specific product categories, such as diapers, food, cat or dog food, where new customers exhibit learning or by choosing product categories with new brand introductions.

of consumer i . This implies that consumer i 's quality beliefs about the unchosen brands do not evolve over time; consequently, for such a consumer, we cannot cleanly distinguish the prior means and variances from the steady-state qualities for these brands. Given these limitations of scanner panel data, researchers have no choice but to impose some homogeneity restrictions on $\nu_{Q_{ij},0}$ and $\sigma_{Q_{ij},0}^2$ to achieve identification. Typically, researchers make assumptions such as setting $\nu_{Q_{ij},0} = \bar{\nu}_j$ and $\sigma_{Q_{ij},0}^2 = \bar{\sigma}_{Q_0}^2$. Furthermore, one assumes that the panel data are sufficiently long so as to allow one to identify Q_{ij} from the individual-level choice share near the end of the sample and $\bar{\nu}_j$ from the difference of choice share between the early and late periods of the sample. Finally, one argues for the identification $\bar{\sigma}_{Q_0}^2$ by relying on the evolution patterns of choice shares over the sample period.

These identification problems are well recognized, and various authors have attempted to mitigate them by making innovative use of available data coupled with adjustments to model structure. Our identification strategy, which is detailed in the next section, is to use individual-level survey data to calibrate the prior constructs, which then will help identify both cross-sectional heterogeneity and the scope and nature of individual learning.

3. Data

In this section we describe the data used in our application. The data consist of typical scanner panel data augmented by a matched survey of the panelists' preferences toward the brands. After presenting the data, we discuss how the survey information is integrated into the model and, along with other assumptions, how it helps identify key constructs pertaining to the Bayesian learning process.

3.1. Data Description

The empirical analysis in this study uses a unique data set on toothpaste choices and preferences obtained from IRI. The scanner panel data contain individual-level choice data over time, along with price and promotion information for the brands within the toothpaste category. Two marketing mix variables, *price* and *in-store display*, are available in this data set. *Price* is measured as the shelf price, inclusive of any temporary price discount. *In-store display* is measured as a scale index ranging from zero to one, which represents the intensity of display activity for a particular brand and time in the relevant store.

A unique feature of the data is that survey information pertaining to *liking* (i.e., how much each respondent likes each brand) and *familiarity* (i.e., how familiar each respondent is with each brand) is available in addition to the standard scanner panel data.

Both *liking* and *familiarity* are rated on a scale from 1 (low) to 7 (high). This stated preference information is pertinent to our identification of individual learning because it was collected from the same individuals we have scanner data on and just before the start of the observation period. It is this additional survey information that will allow us to tease out cross-sectional variation and better initialize time-varying components in the learning process.

The data set comprises a random sample of 673 households dispersed across the United States. Brand choices among seven national brands in the toothpaste category—Aim, Arm & Hammer, Aquafresh, Colgate, Crest, Mentadent, and Pepsodent—were tracked for one year. These seven brands totaled 86% of U.S. category sales at the time. From 673 households, we use only those who made at least four purchases over the study period. This yields a sample of 354 households, making a total of 2,501 purchases in the category.

Table 1 presents basic descriptive statistics related to both survey and scanner data. The two large market share brands, Colgate and Crest, are not the highest-priced brands but, on average, rated high in both liking and familiarity. When compared with Colgate, Crest is priced lower, and displayed less frequently, but rated higher in both liking and familiarity. Furthermore, these two market leaders are repeatedly purchased more often than other brands except Mentadent. The two small market share brands, Aim and Pepsodent, are among the lowest-priced brands and, on average, rated low in both liking and familiarity. The medium market share brands—Aquafresh, Mentadent, and Arm & Hammer—generally rank in the middle in terms of price, display, and survey ratings. There are a couple of noticeable exceptions. Arm & Hammer is the least frequently displayed brand, whereas Mentadent is the highest-priced brand and among the most repeatedly purchased brands. Table 1 also presents demographic information of the sample pertaining to family size and household income. The average family size is about three, and the average household income belongs to the bracket between \$45,000 and \$55,000. These numbers closely match with figures from the 2000 U.S. Census.⁵

3.2. The Information Content of Survey Data

To begin with, we seek to address the issue of whether *liking* and *familiarity* are indeed separate constructs. Figure 1 provides jittered scatterplots of the two constructs for each brand along with their marginal histograms. The correlation between the *liking* and

⁵ According to the 2000 U.S. Census, the average family size is 3.14, and the average household income is \$51,855.

Table 1 Descriptive Statistics for Toothpaste Data

| Variable | Brand | | | | | | |
|------------------------------------|--------------|--------|-------------|---------|--------|-----------|-----------|
| | Arm & Hammer | Aim | Aquafresh | Colgate | Crest | Mentadent | Pepsodent |
| <i>Market share</i> | 0.0704 | 0.0260 | 0.1523 | 0.3179 | 0.3123 | 0.1052 | 0.0160 |
| <i>Repeat purchase probability</i> | 0.4650 | 0.3750 | 0.5138 | 0.5725 | 0.5759 | 0.5764 | 0.2813 |
| <i>Price</i> | | | | | | | |
| Mean | 2.7653 | 1.4250 | 2.3676 | 2.5343 | 2.4377 | 3.5522 | 1.3265 |
| Std. dev. | 0.4860 | 0.4859 | 0.4561 | 0.3915 | 0.4309 | 0.5319 | 0.2332 |
| <i>Display</i> | | | | | | | |
| Mean | 0.0440 | 0.0674 | 0.1141 | 0.2751 | 0.1083 | 0.0980 | 0.0428 |
| Std. dev. | 0.1483 | 0.1747 | 0.1470 | 0.2568 | 0.1435 | 0.1984 | 0.1385 |
| <i>Liking</i> | | | | | | | |
| Mean | 3.4492 | 3.3164 | 4.2006 | 5.4463 | 5.9802 | 4.2486 | 2.8757 |
| Std. dev. | 1.9611 | 1.8194 | 1.9778 | 1.7744 | 1.5507 | 2.1437 | 1.6903 |
| <i>Familiarity</i> | | | | | | | |
| Mean | 4.4011 | 4.3446 | 5.3418 | 6.1045 | 6.2994 | 5.0339 | 3.9689 |
| Std. dev. | 1.7773 | 1.7540 | 1.6488 | 1.3706 | 1.3635 | 1.9305 | 1.8381 |
| | | | | | | | |
| | Demographics | | Family size | | Income | | |
| | Mean | | 3.0198 | | 8.0672 | | |
| | Std. dev. | | 1.2872 | | 2.3568 | | |

Notes. Family size is the number of individuals in the household. Income is measured using categories (1 = less than \$10K, 2 = \$10–\$12K, 3 = \$12–\$15K, 4 = \$15–\$25K, 5 = \$25–\$35K, 6 = \$35–\$45K, 7 = \$45–\$55K, 8 = \$55–\$65K, 9 = \$65–\$75K, 10 = \$75–\$100K, 11 = greater than \$100K).

familiarity ratings hover around the 45%–50% range, and the proportion of consumers providing identical ratings for *familiarity* and *liking* range from 25% (Pepsodent) to 61% (Crest). The spread in the data reveals that consumers are clearly heterogeneous across the measures and also that the covariance between the two is not perfect. Finally, in the reduced-form models of brand choice we discuss next, the effects of the two constructs were separately identified and significant. These tests lead us to conclude that the two measures are indeed separate.

The incorporation of survey data into the model is valuable to the extent it helps explain individual choices. We investigate this further by running a number of reduced-form models that use only the survey data to explain brand choices.⁶ First, we run a simple multinomial logit that seeks to model choices as a function of the survey data measures. The estimated within-sample hit rate was approximately 53.5%, suggesting that cross-sectional heterogeneity across consumers can explain a large proportion of brand choices even without the inclusion of temporal covariates or learning. This model uses both the familiarity and liking scores to predict choices. Excluding the familiarity construct lowered the hit rate to 51%, whereas excluding liking reduces the hit rate to 38%.

⁶ For brevity we have not included the actual results from all reduced-form models we ran but are providing a short qualitative description of the results obtained. Detailed results are available from the authors upon request.

Our informal examination of the survey data underlines its potential role in identifying key effects in choice models and, in particular, structural models incorporating state dependence and learning. In what follows we now examine patterns in the data that might (or might not) support the presence of learning.

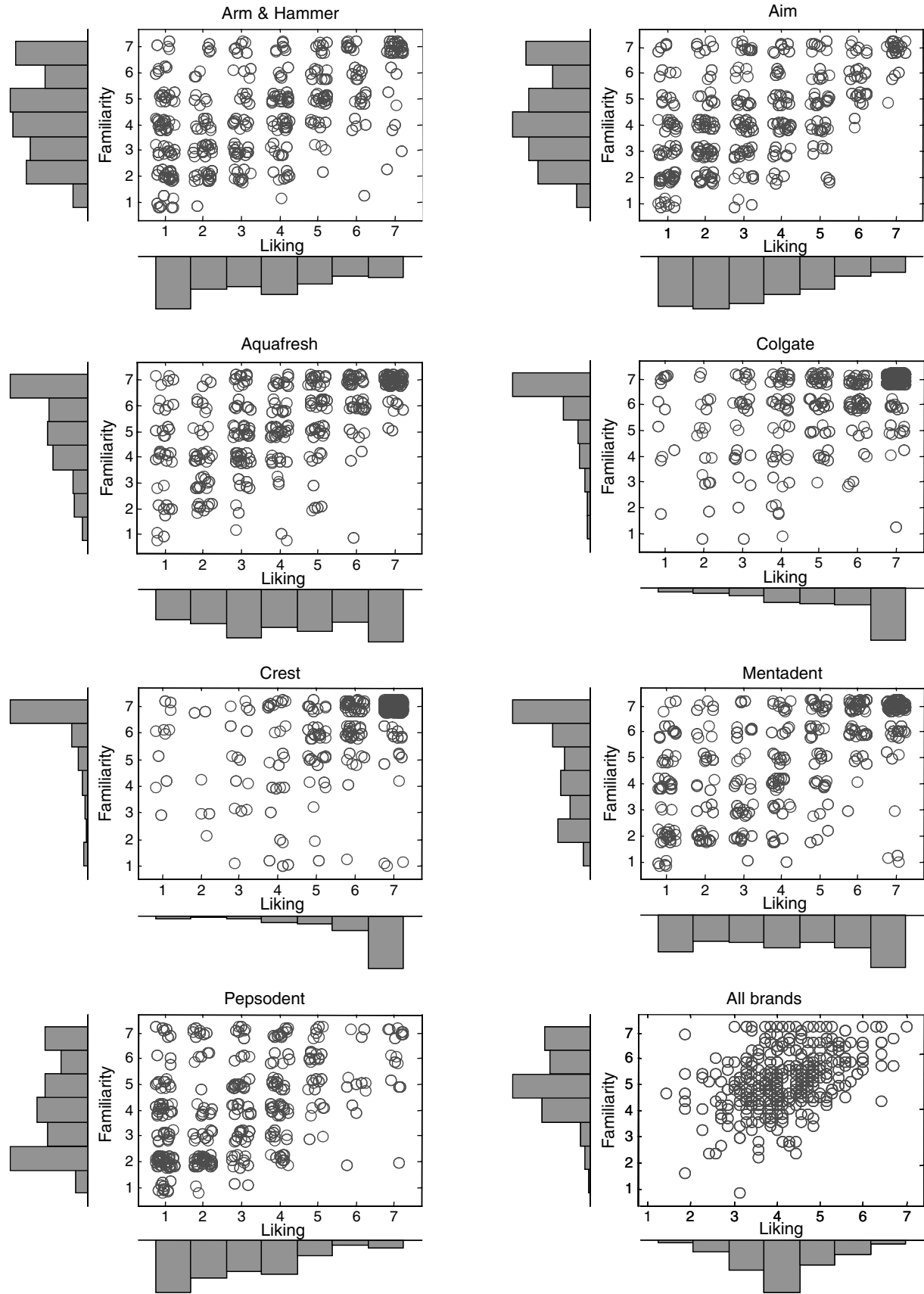
3.3. Where Is the Learning? Model-Free (Lack of) Evidence

Before we embark on a fully structural approach to the problem, we undertake an extensive model-free investigation of the presence of learning in our data set. Given the availability of survey data, we should be able to discern whether consumers are learning without recourse to a full model. To do so we focus our attention on three empirical indicators of learning.

3.3.1. Temporal Effects of Survey Data. Learning, on the part of the consumers, implies that preferences for the brands are temporally varying, and such variation can be tested for. In particular, if consumers are actively learning, choices documented in “early” observations should be significantly affected by liking and familiarity measures, whereas “later” observations should not.⁷ To investigate this we ran separate simple MNL choice models (with and without unobserved heterogeneity) based on two distinct samples created for early and late observations. To ensure robustness we defined early (and late) in various ways, including first and last observations for

⁷ We thank an anonymous reviewer for suggesting this test and other model-free tests.

Figure 1 Scatterplots of Liking and Familiarity with Marginal Histograms



each household, and extended this to first and last “few” observations (two to three observations). In both cases, we found that liking and familiarity are significant predictors of choice. Furthermore, the differences in the effect of these measures were not different from zero in a statistically significant way. This leads us to conclude that preferences are stable and that aggregate patterns in the data do not support a learning hypothesis.

3.3.2. Aggregate Patterns in Market Share. Shifts in market share and changes in variability of market share are often cited as evidence of learning. We ran a number of tests to ascertain whether there were shifts in market share over the sample period conditional on the familiarity levels of the brands. In particular, we examined the movement in market share among those households that indicated low familiarity with brands. If there was indeed learning at play, we should see market shares of these brands changing over time. Although we found no evidence of shifts in market share of the major brands (Crest, Colgate, and Aquafresh), we did find some evidence that market shares were moving over the sample period for niche brands (Arm & Hammer and Mentadent). We note that these shifts could also be explained by movements in prices or promotions. These alternative explanations will be examined when we implement the full model in what follows.

3.3.3. Higher-Order State Dependence. Learning can be thought of as higher (than first)-order state dependence. Clearly, an infinite-order state dependence model cannot be empirically distinguished from learning. To test whether the data exhibit such higher-order patterns, we ran simple choice models with increasing lagged choice indicators. We found that lagged indicators up to the fourth order were significant predictors of current choice. This leads us to conclude that a model with a Bayesian learning component would obtain traction (if higher-order state dependence were ignored, as it usually is in the literature). To check whether this was indeed learning, we included our survey measures into the model with higher-order state dependence. We find that most lagged choice indicators (except of the first order) are insignificant in the presence of the survey data. This again leads us to suspect that any learning in these data may be spurious.

Overall, the results from our model-free analyses suggest that the survey data have significant relevance as a measure of cross-sectional heterogeneity, as a proxy for consumers’ beliefs, and in the data’s ability to disentangle learning from preferences. In the next section, we discuss our empirical implementation and provide details about how the survey data are used to calibrate the levels and uncertainty in consumers’ prior beliefs.

4. Empirical Implementation

To estimate the parameters of our model, we construct a Markov chain Monte Carlo (MCMC) scheme that provides us draws from the stationary joint posterior density of the parameters. In what follows we describe the procedure in brief and relegate details to the appendix.

4.1. Basic Specification

Recall that our basic specification relies on consumers maximizing expected utility of the form

$$U_{ij,t}^E = Q_{ij} + \nu_{Q_{ij},t-1} + X_{ij,t}\beta_i + \varepsilon_{ij,t}^U, \quad (12)$$

with the perception bias at time t being denoted by

$$\nu_{Q_{ij},t-1} = \left[\frac{\sigma_{Q_{ij}}^2}{\sigma_{Q_{ij},0}^2} + \sum_{\tau=1}^{t-1} y_{ij,\tau} \right]^{-1} \left[\frac{\sigma_{Q_{ij}}^2}{\sigma_{Q_{ij},0}^2} \nu_{Q_{ij},0} + \sum_{\tau=1}^{t-1} y_{ij,\tau} \eta_{ij,\tau}^S \right]. \quad (13)$$

In our specification, the vector $X_{ij,t}$ consists of prices and display levels for each brand. Finally, as mentioned before, the $\varepsilon_{ij,t}^U$ are assumed to be distributed *iid* extreme value Type I.

4.2. Identification Using Survey Information

As we mentioned earlier, the identification of learning models is a nontrivial matter. In the traditional, homogeneous learning model, identification restrictions are imposed to achieve identification (see, e.g., Erdem and Keane 1996). More recently, Narayanan and Manchanda (2009) exploit the variation in the patterns of evolution in prescriptions across multiple drugs to identify heterogeneous learning. In both cases, parametric identification assumptions are required on the agent’s prior beliefs to say something about identification. In this paper, we have access to survey data that allows us to inform the model about heterogeneity in prior beliefs without resorting to a purely distributional assumption.

The survey component in our data provides additional information such as liking and familiarity for each brand. Define $S_{ij} = \{\text{LIK}_{ij}, \text{FAM}_{ij}\}$, where S_{ij} is consumer i ’s survey data for brand j , LIK_{ij} is consumer i ’s 1-to-7-point liking measure for brand j , and FAM_{ij} is consumer i ’s 1-to-7-point familiarity measure for brand j .⁸ In both constructs, 1 implies less and

⁸ The results in the sequel are based on using the survey data as is. We also experimented with alternative standardizations of the data and found similar results. The key identification in our framework stems from the variation in the survey responses across individual households, and therefore our results are fairly robust to the way this information is used. Also note that because the impact that liking has on quality is heterogeneous, any scaling distortions will be subsumed into the relevant parameter. Again, this results in our basic findings being robust to the manner in which the survey data are coded.

7 implies more. For example, a 7 on *familiarity* would imply that the consumer is very familiar with that particular brand.

Because this survey information is collected prior to the choices being observed, *liking* and *familiarity* are likely to contain relevant information about the mean and variance of quality perception at the initial period. We exploit this analogy by making the consumers' prior constructs a function of the survey data as follows:

$$\nu_{Q_{ij,0}} = \bar{\nu}_j + \phi_i \left(\text{LIK}_{ij} - \frac{1}{N} \sum_{i=1}^N \text{LIK}_{ij} \right) \quad \text{and} \quad (14)$$

$$\frac{1}{\sigma_{Q_{ij,0}}^2} = \exp(\bar{\kappa} + \delta_i \text{FAM}_{ij}), \quad (15)$$

where the bar notation over the parameters indicates that they are common across consumers. The *liking* and *familiarity* measures pin down a given consumer's quality beliefs and uncertainty at the time the survey was conducted. Because our data pertain only to choices made after the survey data are collected, we are in essence "initializing" our prior constructs at this point. The variation in *liking* allows us to identify heterogeneity in the perception bias, whereas *familiarity* helps identify variation across consumers in prior uncertainty. Together, they allow us to identify a model that allows for heterogeneous priors and, consequently, heterogeneous learning.

When survey information is not available (or is ignored, as in some models we estimate) we set $\phi_i = 0$ and $\delta_i = 0$. Consequently, $\nu_{Q_{ij,0}} = \bar{\nu}_j$ and $1/\sigma_{Q_{ij,0}}^2 = \exp(\bar{\kappa})$. In this case, the initial perception bias is pooled across consumers, and the initial perception variance is pooled across both consumers and brands. These are standard identification restrictions that are needed in the context of Bayesian learning models applied to scanner panel data (see, e.g., Erdem and Keane 1996, Mehta et al. 2003).⁹ Note that, in general, homogeneity restrictions on $\nu_{Q_{ij,0}}$ and $\sigma_{Q_{ij,0}}^2$ do not necessarily imply homogeneous learning. As is evident from Equation (8), the mean quality perceptions are still heterogeneous on account of heterogeneous (yet time-invariant) true mean qualities and the observed sequence of individual-level brand choices. The initial market shares of the toothpaste brands help identify the initial perception bias pooled across consumers. On the other hand, the initial perception variance pooled across both consumers and brands is identified from the evolution patterns of consumer choice behavior for all brands in the market and its relationship with quality signals from consumption experience. Finally, not

all brand-specific $\bar{\nu}_j$ s are identified, and therefore one of them needs to be locationally fixed (say, $\bar{\nu}_K = 0$).¹⁰ This is the last condition to render the parameters in our Bayesian learning model identified.

4.3. MCMC Estimation Scheme

Complete details of the MCMC procedure are in the appendix. The basic algorithm is similar to those used in the literature and involves iteratively sampling parameter blocks from their conditional posterior densities.¹¹ We ran 25,000 iterations after 25,000 burn-in iterations, and we thinned the chain by retaining every fifth draw to reduce autocorrelation, leaving us with 5,000 draws that were then used for inference. Figure 2 presents trace plots for the learning parameters with and without survey data. Visual inspection suggests that the above burn-in period is adequate and that the chains converge. Other parameters exhibit similar patterns and are omitted for the sake of brevity.

5. Results and Empirical Findings

In this section we report our results and discuss empirical findings. These include estimates from the Bayesian learning model with and without the individual-level survey information on familiarity and preferences of the brands. We focus on four areas of interest that pertain to our earlier discussion: (i) model fit, (ii) parameter estimates, (iii) magnitude of learning, and (iv) individual-level insights.

5.1. Model Fit

Table 2 provides the fit statistics for the Bayesian learning model and other competing models. The model fit statistics presented are log-marginal likelihoods computed using the harmonic mean approach of Newton and Raftery (1994), and in all cases the significance of the fit improvement is interpreted based on the criteria proposed by Kass and Raftery (1995).

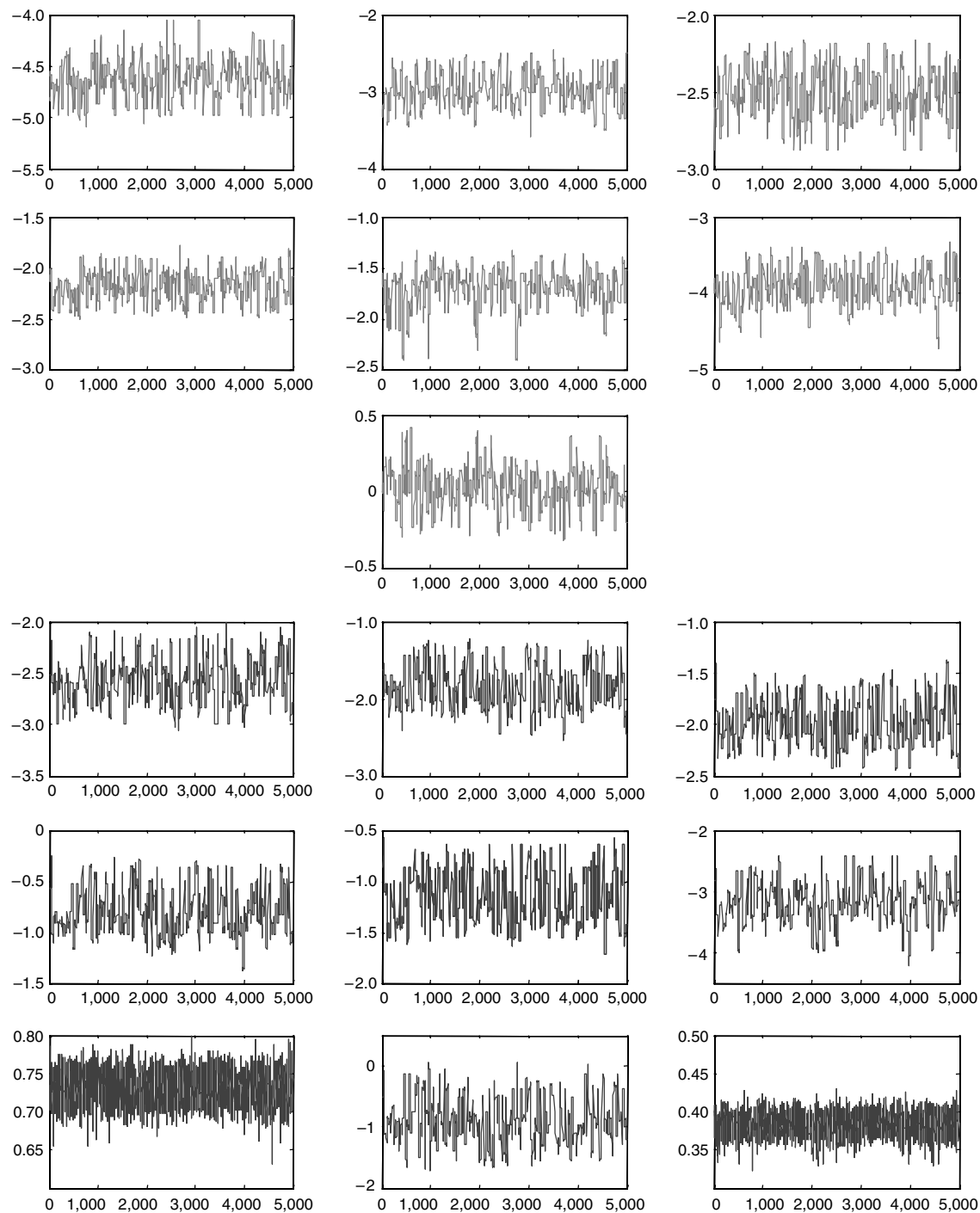
The results presented in Table 2 show that either with or without the survey data, the Bayesian learning model provides a better fit than do any of its competing models (in which state dependence is either not allowed or specified with a different functional form). The differences between the Bayesian learning model log-marginal density and its best-fitting

¹⁰ This K need not be the same as the J that sets $Q_{Jj} = 0$. Also note that in cases where households only buy a single brand in the sample, identification of learning is partially parameteric and relies on the survey data (for initial conditions) and on Bayesian shrinkage for the true mean qualities of other brands.

¹¹ We checked the performance of our estimation procedure with simulated data. The simulation results show that the proposed MCMC sampler converges after several thousand iterations and all parameter estimates recover the true values within sampling error. Full details of the simulation results are available from the authors upon request.

⁹ Mehta et al. (2003) use an initialization sample to minimize the negative impact of these homogeneity assumptions.

Figure 2 MCMC Trace Plots for Learning Parameters



competing model can be classified as “very strong.”¹² Taken at face value, these results make a strong case for the inclusion of the Bayesian learning process in traditional brand choice models. However, we note that the better fit stems directly from the additional

flexibility afforded by the learning framework. In particular, learning-based models permit the incorporation of higher-order feedback effects into the model that mechanically increase performance. The extent to which learning is relevant in our application will be discussed in the sections that follow.

Table 2 also showcases the importance of including survey data. Incorporating survey information always improves fit, irrespective of the model being

¹² Kass and Raftery (1995) suggest that $2\ln(\text{BayesFactor})$ be larger than 10 for the evidence to be very strong in favor of the numerator model. The $\ln(\text{BayesFactor})$ in the above is the difference of the log-marginal densities.

Table 2 Log-Marginal Densities of the Estimated Models

| Model | Data | |
|---------|-----------------------|-----------------------|
| | Standard scanner data | Survey augmented data |
| Model 1 | −2,086 | −1,863 |
| Model 2 | −2,066 | −1,850 |
| Model 3 | −1,983 | −1,820 |

Notes. Model 1: Random coefficient logit without state dependence. Model 2: Random coefficient logit with a loyalty variable (last purchase dummy). Model 3: Random coefficient logit with the Bayesian learning process.

considered. In particular, the fit of the Bayesian learning model improves from −1,983 to −1,820 when the survey information is incorporated. This fit improvement offers very strong evidence in favor of incorporating the survey data. Equally striking is the fact

that the fit of the survey augmented random coefficient logit model without state dependence (i.e., zero-order behavior), in which the individual-level preference measures serve only to “shift” the brand-specific constants, has a better fit than does the Bayesian learning model, which allows a high-order effect of state dependence but does not include the survey information (−1,863 versus −1,983). This implies that stated preferences play a larger and more significant role in explaining choices than does the flexibility afforded by learning models.

5.2. Parameter Estimates

Table 3 provides the parameter estimates of the Bayesian learning model (with and without survey data).

Table 3 Parameter Estimates of the Bayesian Learning Model

| Parameter | Standard scanner data (Choice + Marketing mix) | | Survey augmented data (Choice + Marketing mix + Survey) | |
|--|---|----------------------------|--|----------------------------|
| | Posterior mean | Heterogeneity ^a | Posterior mean | Heterogeneity ^a |
| True mean quality | | | | |
| <i>Arm & Hammer</i> | 10.0911 (9.9696, 10.2133) | 0.4879 (0.3685, 0.6254) | 7.8784 (7.6361, 8.1190) | 0.3618 (0.2671, 0.4423) |
| <i>Aim</i> | 2.7086 (2.5554, 2.8663) | 0.6405 (0.4786, 0.7661) | 1.5366 (1.4073, 1.6578) | 0.4610 (0.2722, 0.6266) |
| <i>Aquafresh</i> | 7.8559 (7.6133, 8.1253) | 0.7972 (0.5203, 0.9752) | 7.2396 (7.0269, 7.4144) | 0.5932 (0.3634, 0.8072) |
| <i>Colgate</i> | 9.3644 (9.2256, 9.5545) | 0.6492 (0.3789, 0.9261) | 8.2139 (7.9863, 8.3593) | 0.4282 (0.2560, 0.6234) |
| <i>Crest</i> | 8.8023 (8.6587, 8.9273) | 0.7165 (0.5484, 1.0392) | 8.3815 (8.1817, 8.6328) | 0.6038 (0.3794, 0.8074) |
| <i>Mentadent</i> | 11.9406 (11.6708, 12.2440) | 0.7473 (0.4750, 0.9930) | 10.8523 (10.5217, 11.1488) | 0.5628 (0.4411, 0.6873) |
| Marketing mix response | | | | |
| <i>Price</i> | −3.7248 (−3.8968, −3.5461) | 1.2648 (1.0784, 1.5095) | −3.4176 (−3.5894, −3.2697) | 0.8011 (0.6862, 0.9105) |
| <i>Display</i> | 0.5192 (0.3717, 0.6821) | 0.7647 (0.5980, 0.9204) | 0.6473 (0.4549, 0.8574) | 0.5866 (0.4030, 0.7786) |
| Initial perception bias | | | | |
| <i>Arm & Hammer</i> | −4.6458 (−4.9804, −4.2964) | | −2.5733 (−2.8897, −2.1705) | |
| <i>Aim</i> | −2.9873 (−3.3147, −2.5927) | | −1.8339 (−2.2444, −1.3390) | |
| <i>Aquafresh</i> | −2.5048 (−2.7724, −2.2344) | | −1.9612 (−2.2965, −1.5967) | |
| <i>Colgate</i> | −2.1824 (−2.4437, −1.9147) | | −0.7872 (−1.0972, −0.3695) | |
| <i>Crest</i> | −1.7016 (−2.0089, −1.4440) | | −1.1369 (−1.5599, −0.6935) | |
| <i>Mentadent</i> | −3.9145 (−4.2794, −3.4891) | | −3.1706 (−3.6424, −2.5867) | |
| <i>Liking</i> − $\overline{\text{Liking}}$ | | | 0.7300 (0.6977, 0.7615) | 0.3433 (0.3215, 0.3656) |
| Log(Initial precision) ^b | | | | |
| <i>Intercept</i> | 0.0221 (−0.2401, 0.2302) | | −0.8785 (−1.5313, −0.2006) | |
| <i>Familiarity</i> | | | 0.3832 (0.3587, 0.4042) | 0.2356 (0.2203, 0.2518) |

Note. Numbers in parentheses indicate a 90% credible set.

^aUnobserved heterogeneity is measured by the posterior mean of the square root of the diagonal element of V_{ψ} (Rossi et al. 1996).

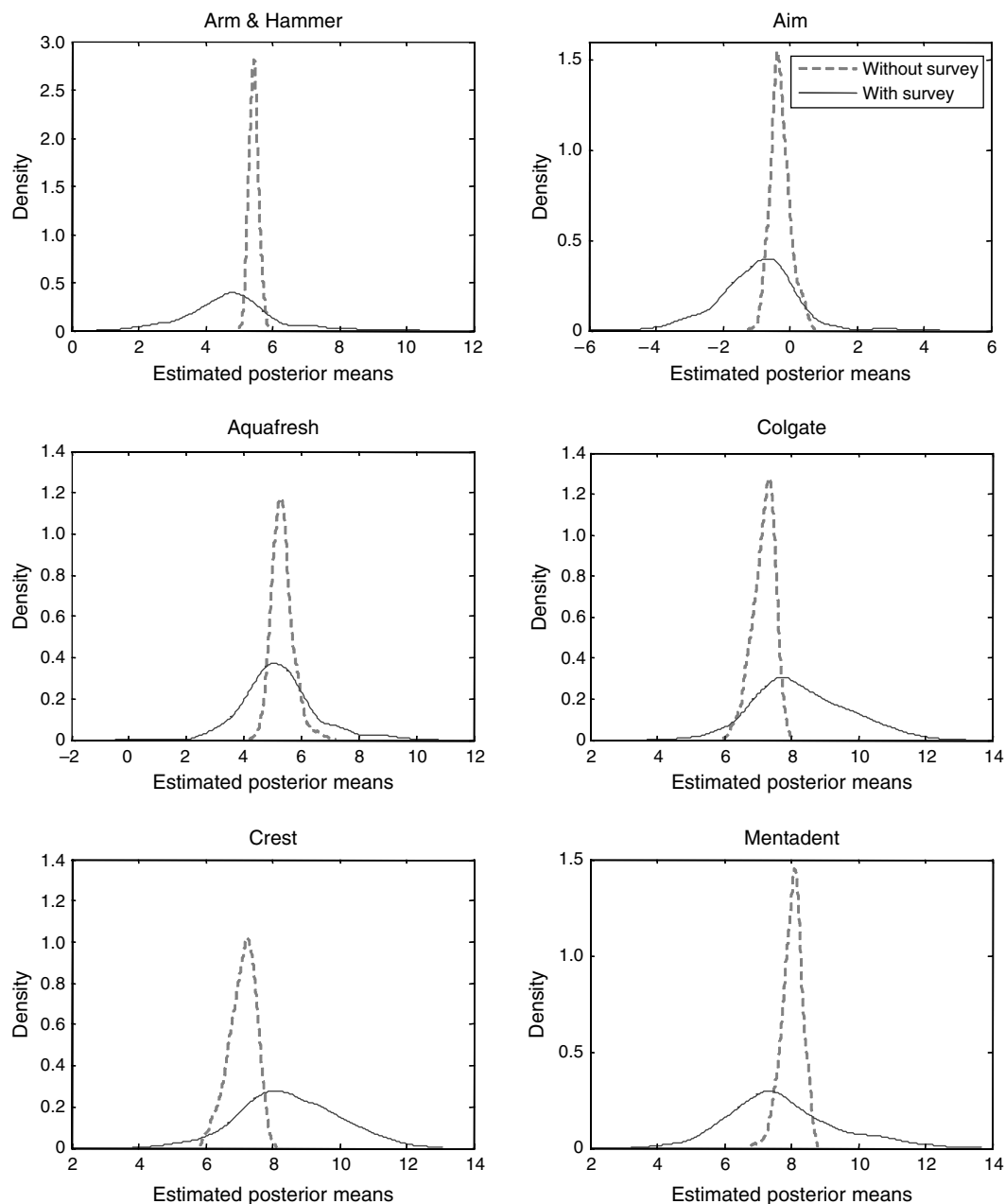
^b $\text{Precision} = 1/\text{Variance of quality perception}$.

5.2.1. Brand-Specific Constants and Qualities.

The inclusion of the stated preference information has a twofold impact on the brand-specific constants, which are interpreted as the true mean qualities in the context of the Bayesian learning model. Pairwise comparisons of these constants (in Table 3) reveal that *both* the mean values and variances are smaller when stated preferences are included. The reduction in the means and in particular in the heterogeneity of the brand-specific constants indicate that stated preferences provide valuable information on the variation in the true mean qualities.

In Figure 3 we plot the estimated individual-specific posterior means of the quality perceptions evaluated at the initial period. This quantity reflects the beliefs of the consumers at the beginning of the sample. The estimated initial mean quality perceptions with survey information are significantly more dispersed for every brand, whereas there is no discernible systematic pattern to the differences when it comes to the location of the density. This suggests that a large proportion of the individual-level variation across consumers is not captured by the brand-specific constants without the aid of the stated

Figure 3 Individual-Specific Posterior Means of Initial Mean Quality Perception



preference information. In the absence of the survey information, some of that variation is carried over to other constructs in the model, such as the heterogeneity in marketing mix effects. We discuss this next.

5.2.2. Sensitivity to Marketing Mix Variables.

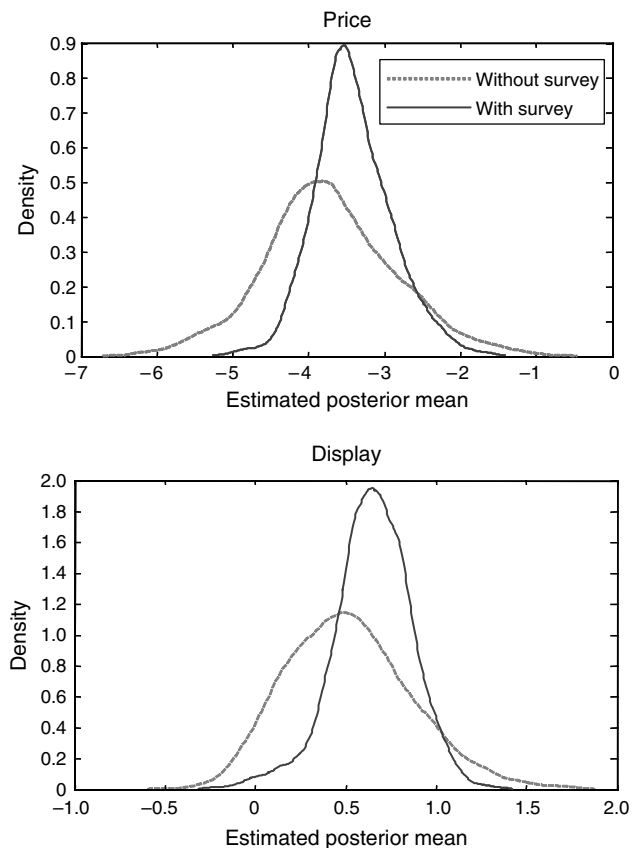
The mean effect and heterogeneity in the sensitivity of marketing mix variables, reported in Table 3, are different when the stated preference information is accounted for. Without survey information consumers are (on average) thought to be more price sensitive and more dispersed in their response to price changes than they actually are. This happens partly because the absence of preference information forces price to account for more than its true effect. It seems that when consumer preferences are “known,” the substitution patterns in choices are well explained, thereby mitigating the need for brand switches to be rationalized by differences in prices. Of course, the full effect of price changes on choices (elasticity) depends not only on the price coefficients but also on the brand-specific constants and true mean qualities. Because for any brand the latter vary across the estimated models, the price elasticities will do so, too. Taken together, the increased heterogeneity in quality beliefs and price sensitivity implies that consumer responsiveness to prices is a lot more varied than traditional models would have us believe. Finally, we note that although aggregate price effects are dampened with the inclusion of stated preferences, there may be individual cases where the effects move in the opposite direction (larger effects with survey data). We will return to these issues in later sections dealing with individual-level insights. Figure 4 depicts the individual-specific posterior means of price and display. Display effects, on average, have a larger mean but, similar to price, exhibit somewhat lower variances when survey data are included.

5.3. Magnitude of Learning

The distribution of individual-level posterior means of the initial perception biases is depicted in Figure 5. Without survey information, initial perception biases for the brands are more negative and less heterogeneous, indicating a larger amount of learning. The extent of learning is determined not only by the initial perception biases but also by variances. Figure 6 presents initial precision, which is defined as inverse of variance of quality perception. Note that in the absence of the survey-based measures, initial perception biases are homogeneous across consumers and initial precision is homogeneous across consumers and brands. These are represented by solid vertical lines in the graphs.

Figure 7 depicts the joint impact of these parameters on aggregate-level learning. There are two significant differences between the two data scenarios.

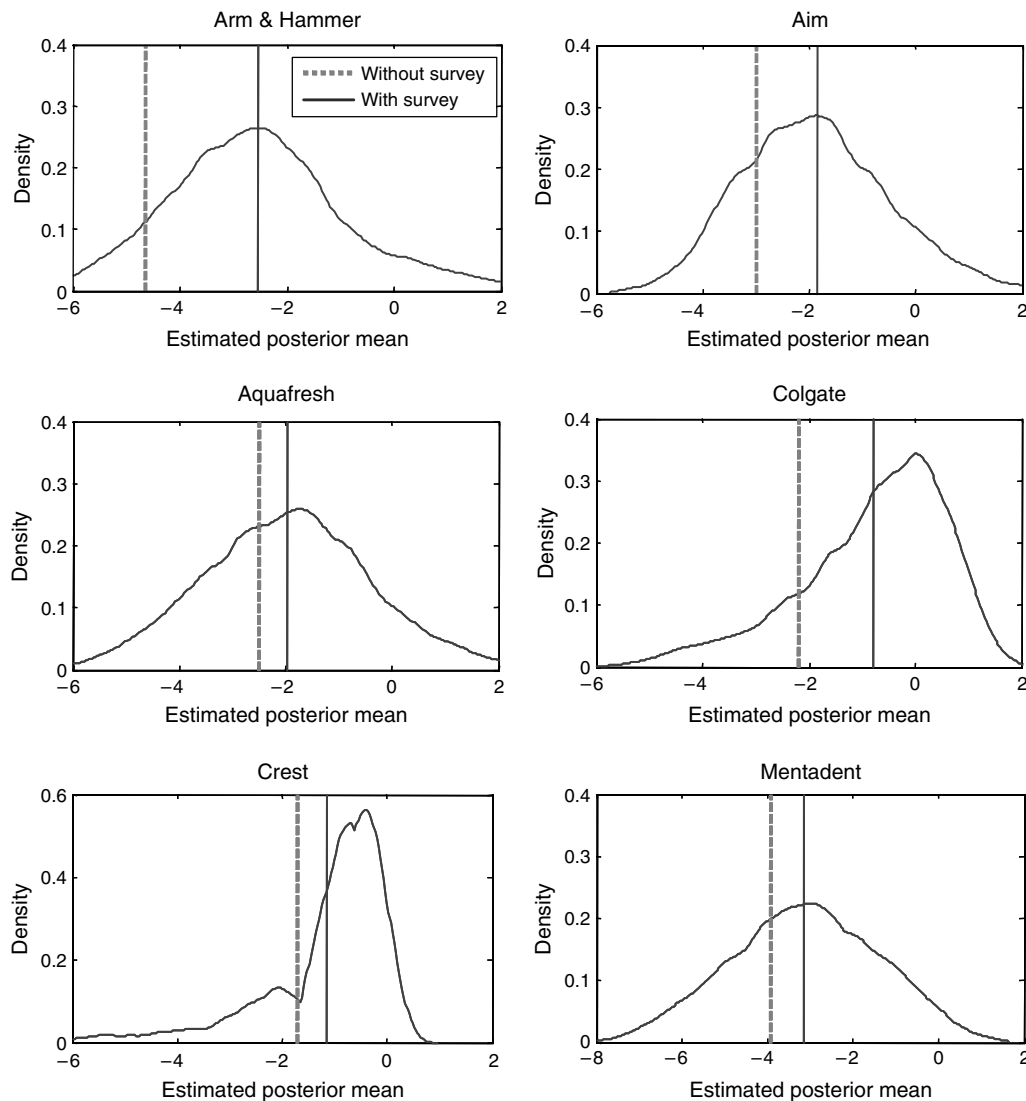
Figure 4 Individual-Specific Posterior Means of Marketing Mix Variables



First, the estimated average learning during the sample period is “lower” when preference information is accounted for. Said differently, the posterior perceived quality levels are much smaller. Second, the rate of learning is also very different. The inclusion of survey information results in consumers updating beliefs at a much “slower” rate. These two effects suggest that the estimated learning patterns without preference information are exaggerated. Note that this pattern of exaggerated learning is true for all brands in the data; however, it is more pronounced for the large-share brands such as Colgate and Crest. In fact, for Colgate and Mentadent, the learning patterns with survey data are not significantly different from a flat line (no learning), whereas for Crest and Aim, the patterns are only weakly different from a no-learning pattern.

Taken together, the utility (constants/marketing mix effects) and learning parameters indicate that without the incorporation of stated preferences, the learning process is forced to proxy for differences across consumers. This reflects the classic confound between state dependence and heterogeneity. By capturing the initial beliefs of individual consumers, we are better able to frame the heterogeneity, which in turn mitigates the need for the learning component

Figure 5 Estimated Posterior Means of Initial Perception Bias



to try to rationalize the unexplained variation. This confound becomes even more stark when we examine individual cases in what follows.

5.4. Individual-Level Insights

Although the patterns discussed in Figure 7 suggest that there is little learning (for most brands) in the data, the aggregate nature of the plots masks the heterogeneous nature of learning. To investigate heterogeneity in learning further, we examine individual-level patterns for the extreme case of Colgate, where the aggregate pattern suggests no learning at all. Figure 8 depicts the learning patterns for each household in our sample for Colgate. There are three important points about this plot: First, including survey data captures heterogeneity in learning. This is depicted in the spread of the curves around the mean curve. Second, in the absence of survey data, the patterns of learning are very similar, whereas

with survey data, the nature of learning is varied. In particular, with survey data, households can *lower* their perceptions about Colgate's quality. Finally, even though the aggregate pattern suggests little learning for Colgate (with survey data), there are households that exhibit significant learning. These insights were echoed for the other brands in our analysis as well.

To further examine heterogeneous learning, we investigate a sample of households at a deeper level. Table 4 depicts four households with different choice patterns facing varied marketing mix environments. For each household, the table also presents liking and preferences. Figure 9 presents the estimated learning patterns for them.

Household #3. This consumer makes five brand choices over the sample period. She chooses Crest four consecutive times and then switches to Colgate on her last shopping trip. The top pair of graphs in

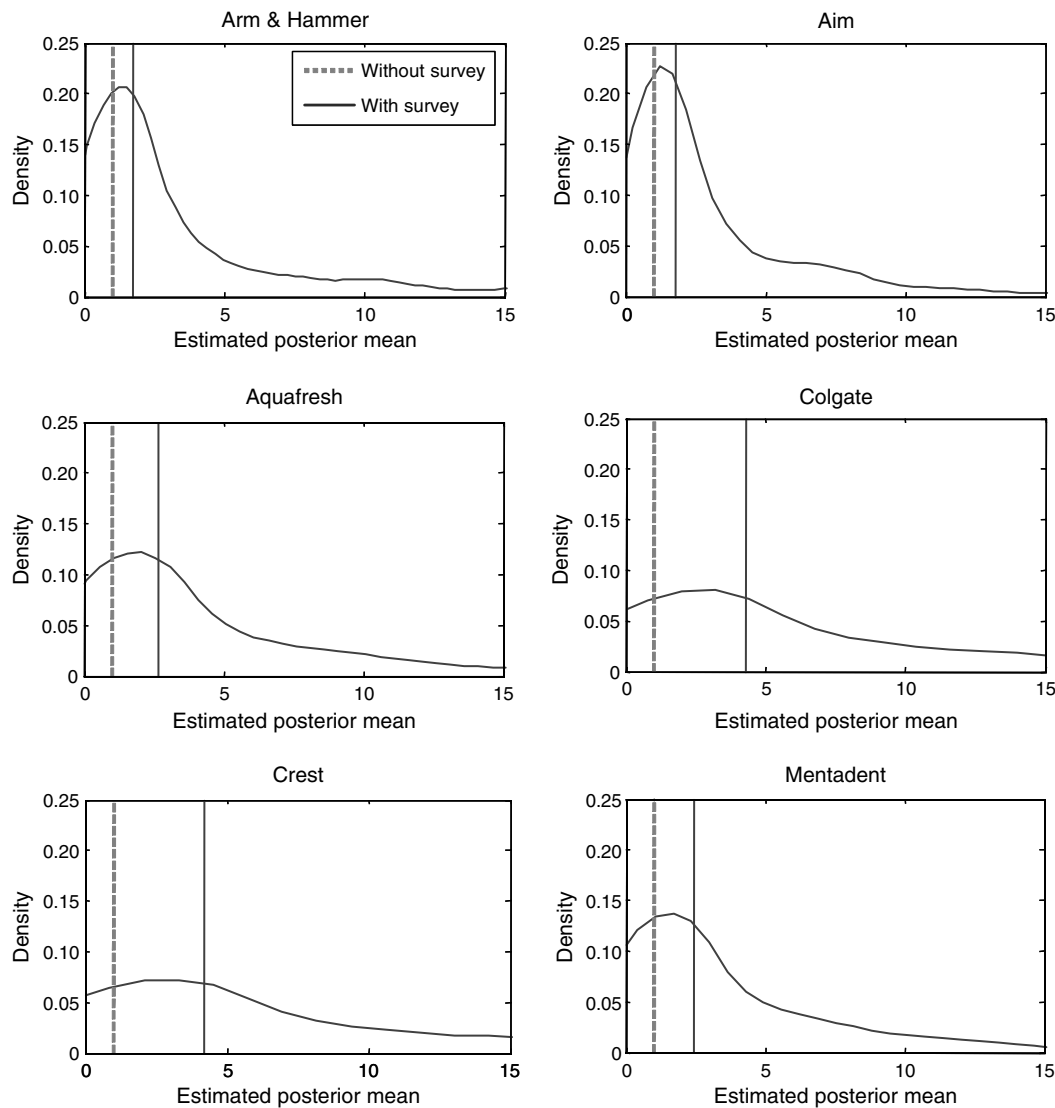
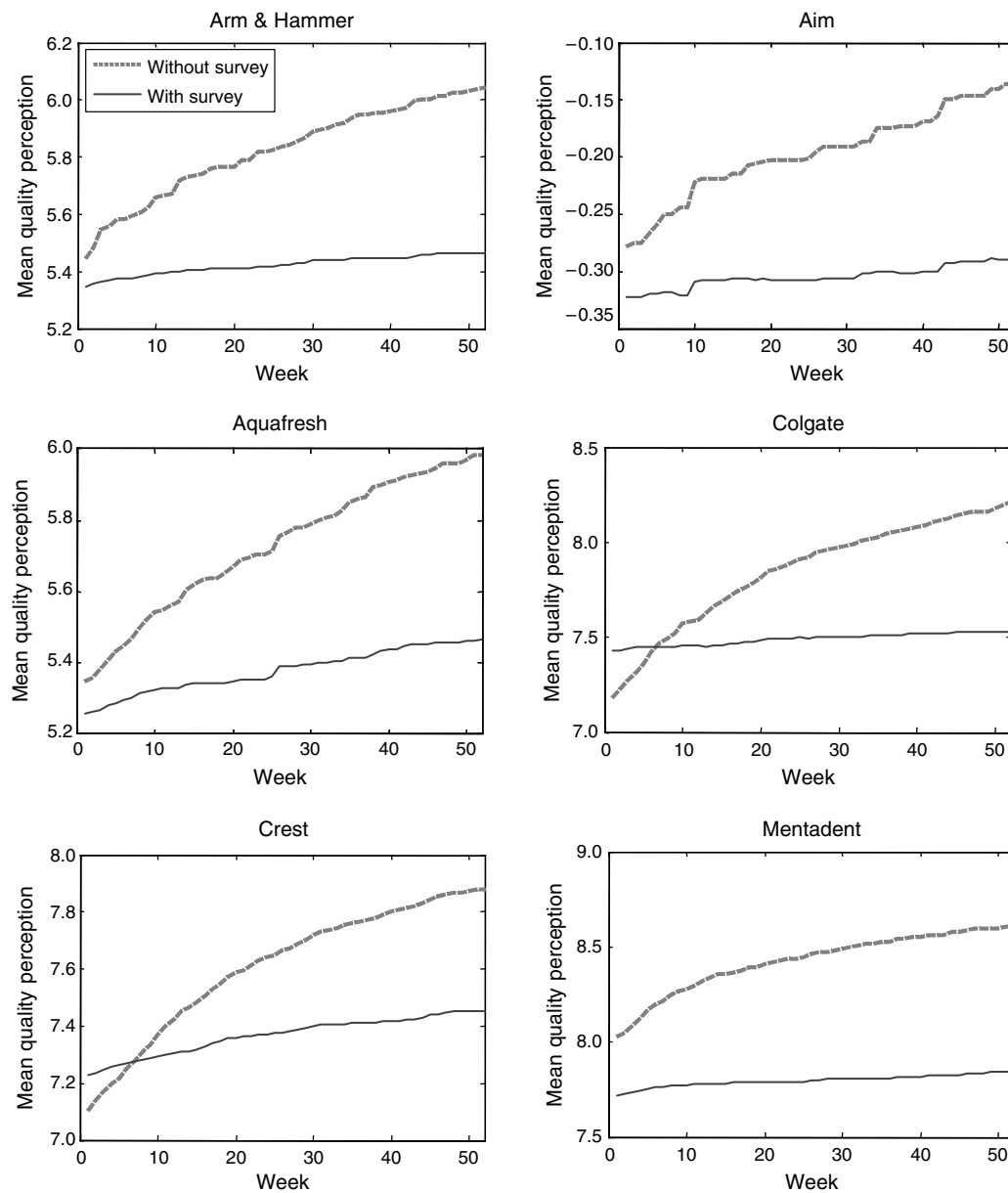
Figure 6 Estimated Posterior Means of Initial Precision

Figure 9 pertains to this household and suggests that she has actively engaged in learning about Crest during the sample period. Her mean quality perception has noticeably increased, and her variance of quality perception has remarkably decreased over successive Crest choices. This preference reinforcement, coupled with uncertainty reduction, indicates active learning about Crest. However, a cursory examination of her stated preferences reveals strong preferences for Crest. Incorporating these data lessens the estimated degree of learning for Crest. This is a clear case where without data, the researcher's learning about consumer preferences is misconstrued as the consumer learning about the brand. As discussed earlier, the survey data also play a role in the consumers' estimated price sensitivity. Because Colgate is also rated favorably, the switch to Colgate on the last purchase occasion does not have to be explained by price differences. Conse-

quently, the estimated price coefficient is -4.13 without survey and -3.50 with survey.

Household #297. This household buys Aquafresh repeatedly and continues to do so even when the price creeps above the mean price. It is only when the price of Aquafresh is significantly above the mean level that she switches over to Arm & Hammer. Her brand choices are unique in that there is no other consumer in the sample who bought Aquafresh six times out of seven. The sample market share of Aquafresh is only 15%. The large disparity between her and the "average" Aquafresh consumer's behavior leads her to be classified as an active learner. A quick examination of the survey data information on liking and familiarity tells a very different story. Aquafresh is not only her most preferred brand but also the one she is most familiar with. Given this information, it is obvious that the consumer buys Aquafresh not because of

Figure 7 Estimated Aggregate-Level Learning

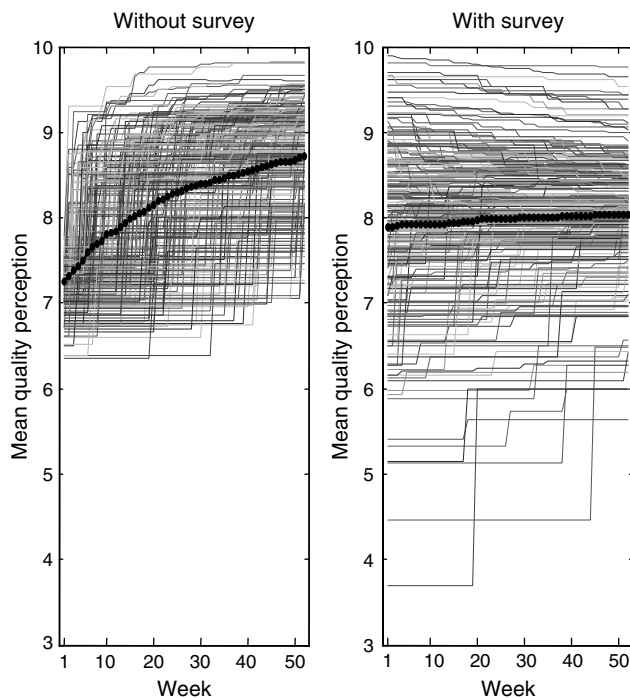


state dependence or learning but simply because she likes the brand. In other words, she is a zero-order type consumer who exhibits no learning whatsoever. Because preferences explain a large proportion of the choice patterns, they also explain why her price coefficient with the survey data is now less negative. The estimated price coefficient is -3.78 without survey and -3.51 with survey.

Households #334 and #55. In contrast to the previously discussed households, no brand switching is observed during the sample period for these two households. One only buys Colgate and the other buys Aquafresh. A quick glance at Table 4, however, reveals that these households differ significantly

in their stated preferences and level of brand familiarity. In a model without the survey information, household #334 appears to be actively learning, but once again the survey data reveal that the Colgate choice can be explained by preferences alone. In contrast, household #55 is identified as learning about Aquafresh even after survey data are included. This happens because the survey data reveal the mean liking of Aquafresh to be 4 and familiarity with Aquafresh to be 3, which are both on the low side of the rating scale. Consequently, the purchase string suggests learning.

Our individual-level analysis uncovered many more examples that offer insights similar to those presented in these examples. For the sake of brevity, we

Figure 8 Estimated Individual-Level Learning for Colgate

have limited ourselves to these cases. We would like to point out that we did attempt to explain the differences in learning across households using demographic covariates. Unfortunately, we had little success in this endeavor, suggesting that differences in learning stem from idiosyncratic differences across households.

6. Discussion

6.1. Implications and Directions

To be clear, our results are based on a single data set and one particular model specification. Although we conjecture that our results are robust to changes in model specification, we also caution the reader that forming generalizations based on our results is not without risk. That said, our results do raise a red flag about the use of learning models in frequently purchased, mature product categories. Although there will be numerous contexts and applications where learning remains an important aspect of consumer behavior, the onus falls on the researcher to document and provide evidence in support of the phenomena.

This paper raises questions about the identification of learning models because different households may have different (and unobservable) initial conditions. Without a strategy for dealing with the heterogeneity in initial conditions, identification of learning is difficult, to say the least. In our application the availability of survey data helps address the problem; however, such data are not universally available. As

such, it would seem that there is no recourse left to researchers wishing to use Bayesian learning models. We do not share this fatalistic attitude. On the contrary, we believe that our results should spur interest in merging varied data sources to learn about consumer preferences. There is already movement in this direction in the marketing and economics disciplines. For example, recent work by Dubé et al. (2009) aims at using a conjoint setting to measure discount factors. This moves us away from the traditional approach to dynamic discrete choice models, which are often identified only from parametric and functional form assumptions. Like them, this paper shows that using *data* to construct consumer beliefs offers new and exciting avenues for research aimed at understanding consumer behavior. Our findings question the blind substitution of structure in place of data and underline the pitfalls of taking identification restrictions for granted. We hope this paper will encourage interest in constructing well-thought-out models where identification is driven more by variation in data than by assumptions.

We recognize that there will be instances where additional data will be unavailable and researchers will need to make strong assumptions to facilitate identification. In such cases, we suggest that they provide evidence as to the robustness of their estimates by perturbing these identification restrictions. In addition, picking categories where learning is easy to justify (diapers, pet food, new products), employing a rich and flexible specification of heterogeneity, and using smart prior initialization and creative identification arguments will all help in convincing the reader that the results obtained are relevant.

On the substantive front, our results highlight a number of interesting issues. We show that a misspecification of the model results in biased estimates for marketing mix effects and for the heterogeneity in them. These clearly have implications for managerial decision making. Our findings also reveal that consumers are heterogeneous not only in the way they react to marketing stimuli but also in terms of the order of their decision process. Although such “process heterogeneity” has been well documented in the literature (see, e.g., Givon and Horsky 1979), there may be reason to allow for such heterogeneity when estimating models such as learning. For example, in the absence of survey data, it might be worthwhile to allow consumers to be endogenously bucketed as “zero-order” or “Bayesian learners” as part of the estimation algorithm. We are currently working on implementing methods in this direction.

Finally, we note that the model implemented in this paper makes a number of assumptions. For example, we assume that consumers are myopic and risk averse. A natural extension would be to relax these

Table 4 Examples of Individual-Level Scanner and Survey Data

| Purchase occasion | Choice | Price | | | | | | |
|-------------------|--------|-------|-------|-------|-------|-------|-------|-------|
| | | AH | AM | AF | CG | CR | MT | PS |
| Household #3 | | | | | | | | |
| 1 | CR | 2.202 | 1.242 | 2.631 | 2.663 | 0.790 | 3.023 | 1.332 |
| 2 | CR | 3.239 | 1.380 | 2.466 | 2.588 | 1.500 | 4.480 | 1.033 |
| 3 | CR | 3.174 | 1.776 | 2.229 | 2.560 | 1.440 | 3.503 | 1.874 |
| 4 | CR | 3.410 | 1.114 | 2.629 | 2.863 | 1.800 | 4.064 | 1.116 |
| 5 | CG | 2.837 | 1.756 | 2.033 | 0.860 | 2.877 | 3.449 | 1.190 |
| Liking | | 2 | 4 | 4 | 7 | 7 | 4 | 2 |
| Familiarity | | 4 | 4 | 5 | 7 | 7 | 3 | 2 |
| Mean price | | 2.765 | 1.425 | 2.368 | 2.534 | 2.438 | 3.552 | 1.327 |
| Household #297 | | | | | | | | |
| 1 | AF | 2.478 | 1.210 | 1.680 | 2.542 | 2.741 | 3.230 | 1.443 |
| 2 | AF | 2.737 | 1.419 | 1.680 | 2.616 | 2.535 | 4.074 | 1.097 |
| 3 | AF | 2.833 | 1.237 | 2.434 | 2.696 | 2.476 | 4.263 | 1.108 |
| 4 | AF | 3.239 | 1.380 | 2.466 | 2.588 | 2.558 | 4.480 | 1.033 |
| 5 | AF | 2.905 | 1.452 | 2.060 | 2.596 | 2.488 | 3.679 | 1.546 |
| 6 | AF | 4.120 | 1.035 | 1.680 | 2.457 | 2.624 | 3.284 | 1.032 |
| 7 | AH | 2.490 | 1.715 | 2.737 | 2.584 | 2.507 | 3.341 | 1.748 |
| Liking | | 4 | 4 | 7 | 4 | 4 | 1 | 4 |
| Familiarity | | 4 | 4 | 7 | 5 | 5 | 2 | 4 |
| Mean price | | 2.765 | 1.425 | 2.368 | 2.534 | 2.438 | 3.552 | 1.327 |
| Household #334 | | | | | | | | |
| 1 | CG | 2.478 | 1.210 | 2.422 | 2.542 | 2.741 | 3.230 | 1.442 |
| 2 | CG | 2.478 | 1.210 | 2.422 | 2.320 | 2.741 | 3.320 | 1.442 |
| 3 | CG | 2.231 | 1.355 | 3.404 | 2.512 | 2.707 | 3.225 | 1.190 |
| 4 | CG | 3.957 | 1.474 | 3.072 | 1.720 | 2.577 | 3.797 | 1.095 |
| 5 | CG | 2.860 | 1.442 | 2.434 | 2.240 | 2.549 | 3.844 | 1.485 |
| 6 | CG | 2.444 | 0.853 | 2.434 | 2.673 | 2.463 | 3.546 | 1.180 |
| 7 | CG | 2.614 | 1.755 | 2.166 | 2.731 | 2.490 | 3.465 | 1.260 |
| 8 | CG | 2.230 | 2.030 | 2.275 | 2.050 | 2.610 | 3.823 | 1.045 |
| 9 | CG | 3.319 | 1.270 | 2.325 | 2.020 | 2.917 | 3.563 | 1.352 |
| 10 | CG | 2.897 | 1.425 | 2.533 | 2.489 | 2.812 | 3.988 | 1.213 |
| Liking | | 1 | 3 | 3 | 7 | 5 | 6 | 3 |
| Familiarity | | 4 | 5 | 7 | 7 | 7 | 7 | 3 |
| Mean price | | 2.765 | 1.425 | 2.368 | 2.534 | 2.438 | 3.552 | 1.327 |
| Household #55 | | | | | | | | |
| 1 | AF | 2.582 | 1.275 | 2.060 | 3.156 | 2.890 | 3.501 | 1.290 |
| 2 | AF | 2.231 | 1.355 | 1.860 | 2.512 | 2.707 | 3.225 | 1.190 |
| 3 | AF | 2.475 | 1.007 | 2.070 | 2.371 | 2.319 | 3.668 | 1.375 |
| 4 | AF | 2.857 | 1.044 | 1.720 | 2.763 | 2.486 | 3.215 | 1.243 |
| Liking | | 2 | 1 | 4 | 2 | 5 | 6 | 3 |
| Familiarity | | 3 | 1 | 3 | 5 | 6 | 7 | 7 |
| Mean price | | 2.765 | 1.425 | 2.368 | 2.534 | 2.438 | 3.552 | 1.327 |

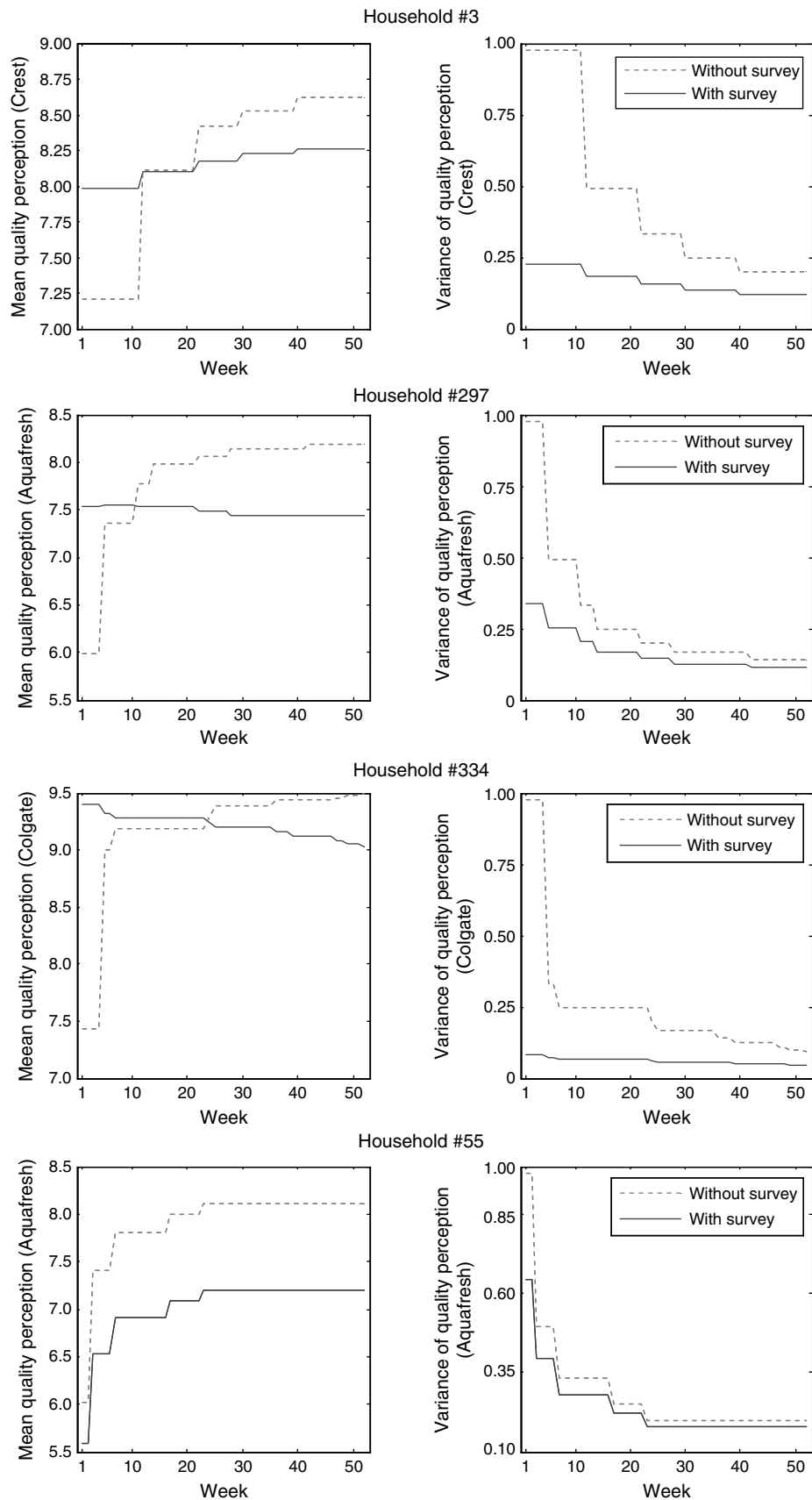
Note. AH, Arm & Hammer; AM, Aim; AF, Aquafresh; CG, Colgate; CR, Crest; MT, Mentadent; PS, Pepsodent.

assumptions and investigate the degree of experimentation that emerges with and without the inclusion of survey data.

7. Summary and Conclusion

Consumers in choosing brands within a product category act intelligently. They use their existing preferences and update those based on their own

consumption experiences. A key problem in the identification of learning models (or state dependence models in general) is that initial conditions are difficult to pin down. Without these initial conditions being known, a clear identification of the degree of learning is all but impossible. In this study we capture consumers' initial beliefs as a function of stated preferences and investigate the impact these data have on the scope, degree, and nature of learning in

Figure 9 Examples of Individual-Level Learning

the sample. Our findings reveal that including stated preferences and familiarity information allows for a better characterization of heterogeneity and reduces the extent of learning.

Our results have implications for both practitioners and scholars. For managers, the findings suggest that consumers' preferences of brands in established categories might be much stronger than extant models would have you believe. This in turn has implications for the effectiveness of pricing and promotional decisions (see, e.g., Dubé et al. 2008, Freimer and Horsky 2008). For scholars our findings reveal new insights into the way learning models are identified and offers avenues for future research in this area.

Acknowledgments

The authors thank the editor, associate editor, and two anonymous reviewers for their comments and suggestions. The paper has also benefitted from comments by seminar participants at the University of Chicago, Stanford University, Purdue University, and the University of Rochester, as well as the *Marketing Science* Conference in Pittsburgh.

Appendix. MCMC Implementation Details

A.1. Overview

The full parameter set of the Bayesian learning model is defined as follows: $\Theta = \cup_i \Theta_i \times \bar{\Phi}$, where $\Theta_i = \{Q_{i1}, \dots, Q_{ij-1}, \beta_i, \phi_i, \delta_i, \eta_{ij,\tau}^S \text{ for } \tau=1, \dots, T_i-1\}$ represents a set of the individual-level parameters, and $\bar{\Phi} = \{\bar{v}_1, \dots, \bar{v}_{J-1}, \bar{\kappa}\}$ stands for a set of the aggregate-level parameters. The expected utility specification is now represented by

$$U_{ij,t}^E = Q_{ij} + \left[\exp(\bar{\kappa} + \delta_i \text{FAM}_{ij}) \left(\bar{v}_j + \phi_i \left(\text{LIK}_{ij} - \frac{1}{N} \sum_{i=1}^N \text{LIK}_{ij} \right) \right) + \sum_{\tau=1}^{t-1} y_{ij,\tau} \eta_{ij,\tau}^S \right] \cdot \left[\exp(\bar{\kappa} + \delta_i \text{FAM}_{ij}) + \sum_{\tau=1}^{t-1} y_{ij,\tau} \right]^{-1} + \beta_i X_{ij,t} + \varepsilon_{ij,t}^U.$$

Given that $\varepsilon_{ij,t}^U$ is iid Type I extreme value distributed, the resulting likelihood function is of multinomial logit form, as given by

$$\mathcal{L}_i(\mathbf{y}_i | \mathbf{X}_i, \mathbf{S}_i; \Theta_i, \bar{\Phi}) = [\mathbf{y}_i | \mathbf{X}_i, \mathbf{S}_i; \Theta_i, \bar{\Phi}] = \prod_{t=1}^{T_i} \prod_{j=1}^J \left(\frac{\exp(\tilde{U}_{ij,t}^E)}{\sum_{q=1}^J \exp(\tilde{U}_{iq,t}^E)} \right)^{y_{ij,t}},$$

where \mathbf{S}_i is individual i 's survey information, $\tilde{U}_{ij,t}^E$ stands for the deterministic part of expected utility, and the bracket notation $[\cdot | \cdot]$ is hereafter used for a generic expression of conditional probability distributions.

To construct an MCMC sampler for the Bayesian learning model presented here, we complete our hierarchical setup by specifying prior distributions for the parameters. For notational simplicity, we further decompose Θ_i into $\Psi_i = \{Q_{i1}, \dots, Q_{ij-1}, \beta_i\}$ and $\Omega_i = \{\phi_i, \delta_i\}$ such that $\Theta_i = \Psi_i \cup \Omega_i$. The former (Ψ_i) represents a set of

the individual-level parameters pertaining to the standard multinomial logit model, whereas the latter (Ω_i) stands for a set of the individual-level parameters unique to the Bayesian learning model. The prior distributions of the model parameters are specified as follows.

1. Individual-level parameters in the standard multinomial logit model, $\Psi_i = \{Q_{i1}, \dots, Q_{ij-1}, \beta_i\}$:

$$[\Psi_i | \bar{\Psi}, V_{\bar{\Psi}}] = \text{MVN}(\bar{\Psi}, V_{\bar{\Psi}}),$$

$$[\bar{\Psi} | p, P] = \text{MVN}(p, P), \quad \text{and} \quad [V_{\bar{\Psi}} | r, R] = \text{InvW}(r, R).$$

2. Individual-level learning parameters, $\Omega_i = \{\phi_i, \delta_i\}$:

$$[\Omega_i | \bar{\Omega}, V_{\bar{\Omega}}] = \text{MVN}(\bar{\Omega}, V_{\bar{\Omega}}),$$

$$[\bar{\Omega} | h, H] = \text{MVN}(h, H), \quad \text{and} \quad [V_{\bar{\Omega}} | g, G] = \text{InvW}(g, G).$$

3. Aggregate-level learning parameters, $\bar{\Phi} = \{\bar{v}_1, \dots, \bar{v}_{J-1}, \bar{\kappa}\}$:

$$[\bar{\Phi} | q_{\bar{\Phi}}, Q_{\bar{\Phi}}] = \text{MVN}(q_{\bar{\Phi}}, Q_{\bar{\Phi}}).$$

4. Signal noises $\{\eta_{ij,\tau}^S\}_{\text{for } \tau=1, \dots, T_i-1}$ in the Bayesian learning model are by design drawn from a standard normal distribution. That is,

$$[\eta_{ij,\tau}^S | \mu_{\eta}, \sigma_{\eta}^2] = \text{N}(\mu_{\eta}, \sigma_{\eta}^2), \quad \text{where } \mu_{\eta} = 0 \text{ and } \sigma_{\eta}^2 = 1.$$

Hyperparameters $p, P, r, R, h, H, g, G, q_{\bar{\Phi}}$, and $Q_{\bar{\Phi}}$ are appropriately chosen to make the corresponding prior distributions diffuse. These prior distributions, coupled with the likelihood function, specify the target posterior distribution from which we need to sample.

Our sampling procedure starts with an initialization of the MCMC sampler. We draw the starting values of $\bar{\Psi}, V_{\bar{\Psi}}, \bar{\Omega}, V_{\bar{\Omega}}, \bar{\Phi}$, and $\{\eta_{ij,\tau}^S\}_{\tau=1}^{t-1}$ from their prior distributions and those of Ψ_i and Ω_i from $\text{MVN}(\bar{\Psi}, V_{\bar{\Psi}})$ and $\text{MVN}(\bar{\Omega}, V_{\bar{\Omega}})$, respectively. Our sampler then cycles through the following steps, each one performed conditional on current values of all other parameters in the model.

Step 1. Update Ψ_i by a Metropolis-Hastings (hereafter, M-H) sampler.

Step 2. Update $\bar{\Psi}$ and $V_{\bar{\Psi}}$ by a Gibbs sampler.

Step 3. Update Ω_i by an M-H sampler.

Step 4. Update $\bar{\Omega}$ and $V_{\bar{\Omega}}$ by a Gibbs sampler.

Step 5. Update $\bar{\Phi}$ by an M-H sampler.

Step 6. Update $\{\eta_{ij,\tau}^S\}_{\tau=1}^{t-1}$ by an M-H sampler.

Sampling procedures in Steps 1 and 2 are now well established in the literature because they are the same as those for a standard random coefficient logit. The subsequent steps involve updating the parameters specific to the Bayesian learning processes. Narayanan and Manchanda (2009) propose an MCMC sampling scheme for a heterogeneous version of the probit-based Bayesian learning model. We adapt their methodology to our logit-based framework by appropriate substitutions of the M-H steps where needed. A noticeable adaptation in our sampling procedure outlined above is that the series of signal noises are sampled independently and updated simultaneously in Step 6, thereby making the chain easier to construct and faster to sample. Full details on the MCMC sampling scheme are presented below.

A.2. Details

Given the prior specifications and the likelihood function of the Bayesian learning model, the joint posterior distribution of all the parameters conditional on the data is proportional to

$$\prod_{i=1}^I \left([y_i | X_i, S_i; \Psi_i, \Omega_i, \bar{\Phi}, \{\eta_{ij,\tau}^S\}_{\tau=1}^{T_i-1}] [\Psi_i | \bar{\Psi}, V_{\bar{\Psi}}] [\Omega_i | \bar{\Omega}, V_{\bar{\Omega}}] \cdot \prod_{\tau=1}^{T_i-1} [\eta_{ij,\tau}^S | \mu_{\eta}, \sigma_{\eta}^2] \right) \times [\bar{\Psi} | p, P] [\bar{\Omega} | h, H] [\bar{\Phi} | q_{\bar{\Phi}}, Q_{\bar{\Phi}}].$$

We here illustrate the MCMC sampling procedure outlined in the estimation section. Following are the details of each of six steps employed to estimate the proposed model in this study.

Step 1. Update $\Psi_i = \{Q_{i1}, \dots, Q_{iT_i-1}, \beta_i\}$ by an M-H sampler.

The full conditional distribution of Ψ_i is

$$[\Psi_i | \text{rest}] \propto [y_i | X_i, S_i; \Psi_i, \Omega_i, \bar{\Phi}, \{\eta_{ij,\tau}^S\}_{\tau=1}^{T_i-1}] [\Psi_i | \bar{\Psi}, V_{\bar{\Psi}}],$$

and we generate a vector of proposal values Ψ'_i using a symmetric random-walk M-H algorithm. The acceptance probability of Ψ'_i is $\min(1, ([\Psi'_i | \text{rest}] / [\Psi_i | \text{rest}]))$. This step is conducted on an individual basis.

Step 2. Update $\bar{\Psi}$ and $V_{\bar{\Psi}}$ by a Gibbs sampler.

Because of the conjugate prior specification for $\bar{\Psi}$ and $V_{\bar{\Psi}}$, their full conditional distributions are

$$[\bar{\Psi} | \text{rest}] = \text{MVN} \left(\frac{V_{\bar{\Psi}}^{-1} \sum_{i=1}^I \Psi_i + P^{-1} p}{V_{\bar{\Psi}}^{-1} I + P^{-1}}, (V_{\bar{\Psi}}^{-1} I + P^{-1})^{-1} \right) \text{ and}$$

$$[V_{\bar{\Psi}} | \text{rest}] = \text{InvW} \left(r + I, \left(R + \sum_{i=1}^I ((\Psi_i - \bar{\Psi})(\Psi_i - \bar{\Psi})') \right)^{-1} \right),$$

from which it is straightforward to sample.

Step 3. Update $\Omega_i = \{\phi_i, \delta_i\}$ by an M-H sampler.

The full conditional distribution of Ω_i is

$$[\Omega_i | \text{rest}] \propto [y_i | X_i, S_i; \Psi_i, \Omega_i, \bar{\Phi}, \{\eta_{ij,\tau}^S\}_{\tau=1}^{T_i-1}] [\Omega_i | \bar{\Omega}, V_{\bar{\Omega}}],$$

and we generate a vector of proposal values Ω'_i using a symmetric random-walk M-H algorithm. The acceptance probability of Ω'_i is $\min(1, ([\Omega'_i | \text{rest}] / [\Omega_i | \text{rest}]))$. This step is conducted on an individual basis.

Step 4. Update $\bar{\Omega}$ and $V_{\bar{\Omega}}$ by a Gibbs sampler.

Because of the conjugate prior specification for $\bar{\Omega}$ and $V_{\bar{\Omega}}$, their full conditional distributions are

$$[\bar{\Omega} | \text{rest}] = \text{MVN} \left(\frac{V_{\bar{\Omega}}^{-1} \sum_{i=1}^I \Omega_i + H^{-1} h}{V_{\bar{\Omega}}^{-1} I + H^{-1}}, (V_{\bar{\Omega}}^{-1} I + H^{-1})^{-1} \right) \text{ and}$$

$$[V_{\bar{\Omega}} | \text{rest}] = \text{InvW} \left(g + I, \left(G + \sum_{i=1}^I ((\Omega_i - \bar{\Omega})(\Omega_i - \bar{\Omega})') \right)^{-1} \right),$$

from which it is straightforward to sample.

Step 5. Update $\bar{\Phi}$ by an M-H sampler.

The full conditional distribution of $\bar{\Phi}$ is

$$[\bar{\Phi} | \text{rest}] \propto \prod_{i=1}^I [y_i | X_i, S_i; \Psi_i, \Omega_i, \bar{\Phi}, \{\eta_{ij,\tau}^S\}_{\tau=1}^{T_i-1}] [\bar{\Phi} | q_{\bar{\Phi}}, Q_{\bar{\Phi}}],$$

and we generate a vector of proposal values $\bar{\Phi}'$ using a symmetric random-walk M-H algorithm. The acceptance probability of $\bar{\Phi}'$ is $\min(1, ([\bar{\Phi}' | \text{rest}] / [\bar{\Phi} | \text{rest}]))$. Notice that this step is conducted for the full sample.

Step 6. Update $\{\eta_{ij,\tau}^S\}_{\tau=1}^{T_i-1}$ by an M-H sampler.

The full conditional distribution of $\{\eta_{ij,\tau}^S\}_{\tau=1}^{T_i-1}$ is

$$[\{\eta_{ij,\tau}^S\}_{\tau=1}^{T_i-1} | \text{rest}] \propto [X_i, S_i; \Psi_i, \Omega_i, \bar{\Phi}, \{\eta_{ij,\tau}^S\}_{\tau=1}^{T_i-1}] \cdot \prod_{\tau=1}^{T_i-1} [\eta_{ij,\tau}^S | \mu_{\eta}, \sigma_{\eta}^2],$$

and we generate proposal values $\{\eta_{ij,\tau}^{S'}\}_{\tau=1}^{T_i-1}$ using an independent M-H algorithm. Their prior density is used to generate independent proposal values. The acceptance probability of $\{\eta_{ij,\tau}^{S'}\}_{\tau=1}^{T_i-1}$ is

$$\min \left(1, \frac{[X_i, S_i; \Psi_i, \Omega_i, \bar{\Phi}, \{\eta_{ij,\tau}^{S'}\}_{\tau=1}^{T_i-1}]}{[X_i, S_i; \Psi_i, \Omega_i, \bar{\Phi}, \{\eta_{ij,\tau}^S\}_{\tau=1}^{T_i-1}]} \right).$$

This step is conducted on an individual basis.

References

- Ackerberg, D. 2003. Advertising, learning, and consumer choice in experienced goods markets: An empirical examination. *Internat. Econom. Rev.* **44** 1007–1040.
- Ben-Akiva, M., T. Morikawa. 1990. Estimation of switching models from revealed preferences and stated intentions. *Transportation Res. Part A* **24** 485–495.
- Camerer, C., T. Ho, J. Chong. 2002. Sophisticated EWA learning and strategic teaching in repeated games. *J. Econom. Theory* **104** 137–188.
- Crawford, G., M. Shum. 2005. Uncertainty and learning in pharmaceutical demand. *Econometrica* **73** 1137–1174.
- Dubé, J.-P., G. Hitsch, P. Jindal. 2009. Estimating durable goods adoption decisions from stated preference data. Working paper, University of Chicago, Chicago.
- Dubé, J.-P., G. Hitsch, P. E. Rossi, M. Vittorino. 2008. Category pricing with state-dependent utility. *Marketing Sci.* **27** 417–429.
- Eckstein, Z., D. Horsky, Y. Raban. 1988. An empirical dynamic model of optimal brand choice. Working paper, Tel Aviv University, Tel Aviv, Israel.
- Erdem, T., M. Keane. 1996. Decision-making under uncertainty: Capturing dynamic brand choice processes in turbulent consumer goods markets. *Marketing Sci.* **15** 1–20.
- Freimer, M., D. Horsky. 2008. Try it, you will like it—Does consumer learning lead to competitive price promotions? *Marketing Sci.* **27** 796–810.
- Givon, M., D. Horsky. 1979. Application of a composite stochastic model of brand choice. *J. Marketing Res.* **16** 258–267.
- Harris, K., M. Keane. 1999. A model of health plan choice: Inferring preferences and perceptions from a combination of revealed preference and attitudinal data. *J. Econometrics* **89** 131–157.
- Heckman, J. J. 1991. Identifying the hand of the past: Distinguishing state dependence from heterogeneity. *Amer. Econom. Rev.* **81** 75–79.
- Hensher, D., M. Bradley. 1993. Using stated response choice data to enrich revealed preference discrete choice models. *Marketing Lett.* **4** 139–151.
- Horsky, D., P. Nelson. 1992. New brand positioning and pricing in an oligopolistic market. *Marketing Sci.* **11** 133–153.
- Horsky, D., Y. Raban. 1988. A Bayesian updating model of dynamic brand choice behavior. Working paper, University of Rochester, Rochester, NY.
- Horsky, D., S. Misra, P. Nelson. 2006. Observed and unobserved heterogeneity in brand-choice models. *Marketing Sci.* **25** 322–335.
- Kass, R., A. Raftery. 1995. Bayes factors. *J. Amer. Statist. Assoc.* **90** 773–795.
- Manski, C. 2004. Measuring expectations. *Econometrica* **72** 1329–1376.

- Mehta, N., S. Rajiv, K. Srinivasan. 2003. Price uncertainty and consumer search: A structural model of consideration set formation. *Marketing Sci.* **22** 58–84.
- Mehta, N., S. Rajiv, K. Srinivasan. 2004. Role of forgetting in memory-based choice decisions: A structural model. *Quant. Marketing Econom.* **2** 107–140.
- Narayanan, S., P. Manchanda. 2009. Heterogeneous learning and the targeting of marketing communications for new products. *Marketing Sci.* **28** 424–441.
- Newton, M., A. Raftery. 1994. Approximate Bayesian inference with the weighted likelihood bootstrap. *J. Roy. Statist. Soc. Ser. B* **56** 3–48.
- Roberts, J., G. Urban. 1988. Modeling multiattribute utility, risk, and belief dynamics for new consumer durable brand choice. *Management Sci.* **34** 167–185.
- Rossi, P. E., R. E. McCulloch, G. M. Allenby. 1996. The value of purchase history data in target marketing. *Marketing Sci.* **15** 321–340.