

DISENTANGLING PREFERENCES, INERTIA AND LEARNING IN BRAND CHOICE MODELS

SANGWOO SHIN

SANJOG MISRA

DAN HORSKY

SIMON SCHOOL OF BUSINESS
UNIVERSITY OF ROCHESTER
ROCHESTER, NY 14627

JUNE 1, 2007

Authors are arranged in reverse alphabetical order.

DISENTANGLING PREFERENCES, INERTIA AND LEARNING IN BRAND CHOICE MODELS

Abstract

The forces that influence a consumer in brand choice: preferences, experiences, and marketing mix have been of great interest to marketing scientists. The vast literature based on static survey data points to consumers using multi-attribute utilities in making a choice. At the same time the vast literature of scanner panel data based studies points to the existence of state dependence (in various functional forms) in brand choices in frequently purchased product markets.

Using a unique dataset 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, and to determine the exact nature of the latter. We propose a hierarchical model in which consumers within the same product category are heterogeneous in the order of the brand choice process and its parameters as well as in their preferences and responsiveness to marketing mix. The proposed model is designed to encompass side by side three different types of consumer experience based behavior: zero-order, inertia and learning. Two sources of state dependence, inertia and learning, are operationalized using a dummy for lagged choice and adopting the Bayesian learning process, respectively. We apply a Reversible Jump MCMC sampling scheme to sample across component processes and a Metropolis-Hastings/Gibbs step within each component process.

Our results are striking and suggest that all three processes exist but that the extent of state dependence is spuriously overestimated in the absence of preference information. Both the inertia and in particular the Bayesian learning components are overstated, and the nature of learning significantly changes. A substantial segment of consumers who do not exhibit state dependence is uncovered and its size increases when preferences are available. Using individual-level outcomes we illustrate why the lack of preference information leads to faulty inferences. Furthermore, we find that sensitivity to marketing mix variables is biased. We also conduct various counterfactual simulations to assess the managerial implications of our findings.

Key Words: Brand Choice, Preferences, State Dependence, Bayesian Learning, Reversible Jump MCMC

1 Introduction

In choosing brands in a frequently purchased product category, consumers are potentially influenced by (i) their prior preferences towards the relevant brands, (ii) the updating of these preferences based on past consumption experiences, and (iii) changes in prices and other marketing mix variables. There is no consensus among scholars who examine consumer brand choice behavior of whether consumption experience matters at all, and if so to what extent – Is there inertia? Are consumers also learning? At the same time, there is no disagreement among scholars that preferences, based on multi-attribute utilities, underlie the choices the consumer makes. However, no direct measure of individual level preferences is available to researchers in typical scanner data which only include information on choices, prices and other marketing mix variables. This “missing” information on individual utilities taints the researcher’s ability to correctly assess the impact of experience and prices on brand choice and may have, in fact, fueled some of the disagreement related to the existence of inertial and learning effects.

To illustrate the key role preferences play, consider the following example: Suppose we observe the choices that Jane makes in the toothpaste category. She chooses Aquafresh six consecutive times and then switches to Arm and Hammer on her last shopping trip. Further examination of the prevailing prices reveals that Jane has been mostly insensitive to price changes in other brands except that she seems to have reacted to a price promotion offered by Arm and Hammer on her last purchase occasion. What can the researcher infer about Jane’s brand choice behavior? Three sharply contrasting behavioral explanations are consistent with Jane’s brand choices during the first six trips: (i) time-invariant preferences (i.e., Jane has repeatedly purchased Aquafresh simply due to Aquafresh being her most preferred brand), (ii) inertia (i.e., Jane, who has happened to buy Aquafresh initially, and has been inertial to her previous choice thereafter) and (iii) learning (i.e., Jane, who was not familiar with Aquafresh tried it initially, and her preference for Aquafresh has been reinforced over successive trials). With typical scanner panel data distinguishing between these competing behavioral explanations is daunting, if not futile, task. This is because, without knowing

Jane's preferences for the brands in question, all of the above three explanations can be rationalized by the data at hand. In other words, state dependence of any form, inertia or consumer learning, is confounded with the researcher's learning (estimation) of Jane's unknown preferences. As a consequence, any inference exhibits a tendency to spuriously overstate the relative importance of state dependence.

What information, then, can researchers collect that would help resolve these issues? The answer, simply, is Jane's preference information. Ideally, if at each purchase occasion, Jane's true preferences for the relevant brands were available, the researcher could pin down the precise underlying behavior that drives Jane's observed choices. Such preference information is tedious and expensive to collect. A second best alternative is to gather Jane's initial preferences at some point before her observed choices. Such information can then be used to resolve the confound between state dependence and preference heterogeneity by allowing these preferences to offer a competing explanation for the observed choice sequence in the estimation procedure. For example, if Jane's preference data reveals that Aquafresh is her most preferred brand and that she is very familiar with the brand, the researcher can immediately rule out a learning based explanation. Sensitivity to marketing mix variables is also better assessed in the presence of such preference information. In Jane's case, a price discount on the last seventh shopping trip seems to have induced her to switch from her most preferred brand (Aquafresh) to a less preferred one (Arm and Hammer). If her preference information were available and indicated that Arm and Hammer was Jane's second most preferred brand, it would imply that she is less price sensitive than if it were her least preferred brand. By similar argument, the effect of other marketing mix elements would also be more cleanly estimated.

Marketer's interest in understanding the underlying behavioral explanations of consumer choice is not a recent phenomena. Consumers' brand switching behavior has been the focus of a large body of research over the last four decades. Early studies were based on consumer diary panel data and due to the inaccuracy of the self reported pricing data only individuals' brand switching information was considered. Various stochastic models were applied to de-

scribe the brand switching patterns. Kuehn (1962) applied the infinite-order linear learning model (LLM) in which all previous purchases impact the current one, while there is geometric decay in the impact of early purchases. Kuehn (1962) assumed in the estimation stage that all consumers have homogeneous process parameters. Frank (1962) demonstrated through simulations, and later Givon and Horsky (1985) provided a formal proof, that heterogeneous zero-order Bernoulli type individuals (whose probabilities of purchasing a specific brand are independent of their previous purchases) will be erroneously identified as infinite-order if homogeneity is imposed.¹ This has led to studies which estimated zero-order heterogeneous models such as Morrison (1966) and the conclusion by several researchers, such as Herniter (1973), Bass (1974), Bass et al. (1984), Uncles, Ehrenberg and Hammond (1995) and Bass and Wind (1995), that consumers are by and large non-learning zero-order types. Heterogeneous first-order Markov models which assume last purchase reinforcement/inertia were investigated by Morrison (1966), Massy, Montgomery and Morrison (1970), and Jeuland (1979). Givon and Horsky (1979) applied a model which allows for both process and parameter heterogeneity. Consumers could be either zero-order, first-order Markov or LLM and within each process could differ in their parameters. They found that while in certain product categories consumers could be all zero-order, or all LLM, in several categories about half were zero-order and half first-order. The above studies used a cross-sectional analysis of short purchase sequences. Blattberg and Sen (1976) employed an individual level analysis based on long purchase strings and also uncovered an even split of zero-order and first-order type individuals within the same categories. At the tail end of these diary panel studies Eckstein, Horsky and Raban (1988) and Horsky and Raban (1988) specified a model of forward looking consumers who updated their utilities in a Bayesian manner. They applied it to data on new brands, short purchase strings and long purchase strings respectively, and reported strong evidence for this type process.

The study of consumers' brand switching behavior received a boost when scanner data with accurate data on pricing, and other marketing mix variables became available. One

¹A similar notion that ignoring consumer heterogeneity will lead to a spurious conclusion of state dependence was advanced by Heckman (1991).

source of confound, that of the possibility that consumers are not inherently stochastic, as argued by Herniter (1973) and Bass (1974), but rather switch due to nonstationarity caused by price changes or price promotions, could be examined. The pioneering work was done by Guadagni and Little (1983) who within the context of a logit model used a GL-type loyalty variable which is akin to the LLM in its formulation. A different type of infinite-order consumer learning based on Bayesian updating was investigated by Erdem and Keane (1996), Ching(2000), Akerberg (2003), and Mehta, Rajiv and Srinivasan (2003). The above set of studies did not however allow for consumer heterogeneity in the learning parameters of the model. The use of unobserved heterogeneity in logit models was advanced by Kamakura and Russell (1989), Chintagunta, Jain and Vilcassim (1991) and Gonul and Srinivasan (1993) who however did not allow for state dependence. These might best be characterized as heterogeneous zero-order logit models. Subsequent authors which allowed for both state dependence and heterogeneity, Roy, Chintagunta and Halder (1996), Keane (1997), Seetharaman and Chintagunta (1998), Seetharaman, Ainslie and Chintagunta (1999), Ailawadi, Gedenk and Neslin (1999) and Seetharaman (2003b), found substantial amount of state dependence.

It is important to note that while the scanner based logit models in contrast to the earlier diary panel data based stochastic models, accounted for price effects they did not allow for the process heterogeneity identified in some of the earlier studies. That is, a mixed population of a segment of heterogeneous zero-order individuals along side a segment of heterogeneous inertial and/or learning individuals was not investigated. Moreover, neither set of studies had information on consumers' brand preferences. In a recent paper Horsky, Misra and Nelson (2006) use preference information within a discrete choice framework. Their specification employs a heterogeneous first-order inertia logit model where the individual preference data is used to "shift" the brand specific constants. As discussed earlier, information on preferences is crucial for the correct measurement of the extent and impact of state dependence and prices.

In the current study we take advantage of the advances offered by both streams of research

and enhance those with information on consumer familiarities and preferences. We make a number of substantive and methodological contributions to the literature. On the methodological front, we introduce and implement a novel logit based composite model of process and parameter heterogeneity which incorporates consumers' familiarities and preferences and the impact of marketing mix variables. This composite model allows three component choice processes (namely, the zero-order, inertia, and Bayesian learning processes) to compete for the best description of the individual level brand choice. The estimation procedure includes a MCMC Reversible Jump step to sample across component processes (e.g., accept/reject moves between any pair of three component processes) and a Metropolis-Hastings/Gibbs step within each component process.

On the substantive front, our findings enhance the current knowledge about the consumer brand choice process. First, we find that process heterogeneity is a critical aspect to describing consumer choices and that all three processes seem to contribute significantly. Second, we demonstrate how the inclusion of preference and familiarity information substantially alters our understanding of the choice process. In particular, the absence of this information overestimates state dependence (learning in particular). In the presence of survey information the heterogeneous zero-order behavior is much more pronounced. The diminishing role of the learning element seems to be due to the elimination of the researcher's learning about the consumer's unknown preferences. Third, our analysis allows us to uncover individual level process heterogeneity parameters, and consequently show the effect that process heterogeneity and preference information have on explaining individual level choice strings such as Jane's. Finally, we find that the inclusion of process heterogeneity and preference information uncovers statistically and managerially significant biases in parameter estimates, such as price sensitivity, and the degree of parameter heterogeneity. Overall, our findings offer the marketing scientist new methods and insights into disentangling the impact of preferences, inertia and learning in consumer's brand choice. As a consequence the marketing manager is also capable of making better marketing mix decisions.

The rest of this paper is organized as follows: In the next section we specify our pro-

posed composite model in which two sources of heterogeneity are accounted for: parameter heterogeneity and process heterogeneity. More specifically, consumers are allowed to be heterogeneous in the order of the brand choice process as well as in their preferences and responsiveness to marketing mix. In the following section we describe our unique data set that 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 familiarity and preferences of the brands is incorporated into the composite model. In particular, our specification of the learning process uses additional parameters that allow the consumer to update initial preferences. We then describe our estimation methodology and follow this with a discussion of our empirical findings. The estimates of the parameters of the composite model which relate to the sizes of its components, state dependence, learning, and marketing mix variables are provided. Comparisons are made with the estimates obtained for special cases of the model, some of which correspond to models previously investigated. We follow with managerial implications of our study. We conduct a series of counterfactual simulations to assess, with and without the survey information, the impact of free sampling, coupon and in-store display. We conclude with a summary.

2 Model Development

The model developed in this section is a composite model of process heterogeneity in which three candidate choice processes compete for the best description of individual level choices. It is a random utility framework extension of the composite heterogeneity model of Jones (1973). Jones (1973) proposed a model in which consumers are allowed to differ in the order of the stochastic process they follow (zero-order Bernoulli, first-order Markov and infinite-order LLM) and also differ in the model parameters within each process. Givon and Horsky (1979) operationalized and estimated the model to investigate the order of the brand choice process in several frequently purchase product markets. Focusing on the learning element Seetharaman (2003a) showed that the multi-brand version LLM is comparable to a random utility model that explicitly accounts for two sources of state dependence: the

lagged choice effect operationalized by the loyalty formulation of Guadagni and Little (1983) and the serial correlation of error terms operationalized by AR(1) process in Allenby and Lenk (1995). Nevertheless, given the growing stream of literature in both marketing and economics which models consumers as Bayesian learners (e.g., Erdem and Keane (1996), Ching (2000), Ackerberg (2003) and Mehta, Rajiv and Srinivasan (2003)), we adopt this type of consumer learning as well. Later in the estimation stage of the paper we will take full advantage of the methodology forwarded by Narayanan and Machanda (2006) who, in the context of pharmaceuticals, were the first to specify and apply a heterogeneous version of the Bayesian learning model.

Given that the Bayesian learning component is the most complex part of our composite model we start with its specification and then proceed to the complete model formulation.

2.1 The Bayesian Quality Learning Process

In the Bayesian learning model, consumers are assumed to learn the true mean quality of brands and update their quality beliefs over successive consumption experiences. More specifically, consumers receive a quality signal after every consumption experience, combine the prior belief with the quality signal, and construct the posterior belief in accordance with Bayes rule. In the Bayesian learning model “learning” is conceptualized as having two distinct effects: *quality perception bias reduction* and *uncertainty reduction*. The first effect stems from the stochastic convergence of a consumer’s quality perception to the true mean quality (*quality perception bias reduction*), while 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. We now present the Bayesian learning framework in detail.

Let $Q_{ij,t}^S$ denote a quality signal consumer i receives by consuming brand j at time t . It is assumed that the quality signal is generated from the following normal distribution:

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

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

Prior to any consumption experience of brand j , 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, $\mu_{Q_{ij,0}}$ and $\sigma_{Q_{ij,0}}^2$ are initial posterior mean and variance of brand j 's quality at time 0. Combining the prior beliefs with the consumption signal allows us to construct the posterior belief at any time $t > 0$. This posterior belief also follows a normal distribution, and is denoted by

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

Since the quality beliefs at any time $t \geq 0$ are normally distributed, they are characterized by mean and variance parameters. In other 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 his/her posterior belief at time $t - 1$ (or prior belief at time t) through a realization of the quality signal in a Bayesian fashion, the posterior mean and variance at time t can be updated in the following recursive manner:

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

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

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

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

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

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, while 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

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

This 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, $\nu_{Q_{ij,t}} = 0$ for $\forall t$), the Bayesian learning process collapses to the zero-order process (i.e., $\mu_{Q_{ij,t}} = Q_{ij}$). This case describes a consumer i who is no longer learning (about brands) since his/her quality perception already converged to the true mean quality and no uncertainty about his/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. For

instance, the popular inertia/purchase reinforcement process is often expressed as $\mu_{Q_{ij,t}} = Q_{ij} + \lambda_i y_{ij,t-1}$. 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), while the Bayesian learning process is a higher than first-order process (which is often referred to as an infinite-order process). More importantly, the nature of state dependence is different. The inertia coefficient λ_i is not varying across brands or over time. In contrast, the effect of learning is heterogeneous across brands and is diminishing over time. As a consequence, these functional differences in modeling state dependence enables us to distinguish one process from the other.

2.2 A Composite Model of Process Heterogeneity

We assume that the brand choice behavior of a given consumer can be described by one of the three candidate processes: zero-order, inertia, or Bayesian learning. We define k as an index for the order of the brand choice process such that the value of k is restricted to be 0, 1, or ∞ .² As is implied by its name, the zero-order process is represented by the case where $k = 0$ while inertia and Bayesian learning correspond to the first- and infinite-order processes, respectively. We also define an individual-specific process indicator w_i such that $w_i = k$ if consumer i follows the k -order brand choice process. Conditional on the value of the individual-specific process indicator w_i , the brand choice processes of consumer i can be represented by either the zero-order, inertia, or Bayesian learning process. More specifically,

$$U_{ij,t}^k = U_{ij,t|w_i=k} = \begin{cases} Q_{ij} + \beta_i X_{ij,t} + \varepsilon_{ij,t}^U & \text{if } w_i = 0 \\ Q_{ij} + \lambda_i y_{ij,t-1} + \beta_i X_{ij,t} + \varepsilon_{ij,t}^U & \text{if } w_i = 1 \\ Q_{ij} + \frac{\frac{\sigma_{Q_{ij}}^2}{\sigma_{Q_{ij,0}}^2} \nu_{Q_{ij,0}} + \sum_{\tau=1}^{t-1} y_{ij,\tau} \eta_{ij,\tau}^S}{\frac{\sigma_{Q_{ij}}^2}{\sigma_{Q_{ij,0}}^2} + \sum_{\tau=1}^{t-1} y_{ij,\tau}} + \beta_i X_{ij,t} + \varepsilon_{ij,t}^U & \text{if } w_i = \infty \end{cases}, \quad (9)$$

²Note that 0, 1 and ∞ are simply process identifiers and are not to be numerically interpreted.

where $U_{ij,t}^k$ (or $U_{ij,t|w_i=k}$) denotes consumer i 's utility of brand j at time t conditional on the value of consumer i 's process (i.e. k). The λ_i parameter captures inertia (the effect of lagged purchase indicators $y_{ij,t-1}$) while the β_i captures the effect of marketing mix variables ($X_{ij,t}$) such as price and display.

In this utility specification the three candidate processes compete for the chance to describe a given individual's brand choice behavior. At first glance our proposed specification looks similar to that of a standard latent class model or a finite mixture formulation (Kamakura and Russel 1989). The key difference, however, is that the utility specification is structurally different across three candidate processes of differing orders. The particular functional form of modeling state dependence turns on and off depending on the value of the individual-specific process indicator. In this sense our proposed model can be better understood as a variant of heterogeneous variable selection model (Gilbride, Allenby, and Brazell 2005).

Under the assumption that stochastic utility components $\varepsilon_{ij,t}^U$ are identically and independently distributed Type-I Extreme Value random variables, the probability that consumer i chooses brand j at time t conditional on the process indicator is of the conditional logit form,

$$\mathcal{P}(y_{ij,t} = 1 | w_i = k, X_{ij,t}; \Theta_i^k) = \frac{\exp(\bar{U}_{ij,t}^k)}{\sum_{q=1}^J \exp(\bar{U}_{iq,t}^k)}, \quad (10)$$

where $\bar{U}_{ij,t}^k$ is a deterministic part of $U_{ij,t}^k$; Θ_i^k is a set of process-specific parameters, with $\Theta_i^0 = (Q_{ij}, \beta_i)$, $\Theta_i^1 = (Q_{ij}, \beta_i, \lambda_i)$, and $\Theta_i^\infty = \left(Q_{ij}, \beta_i, \nu_{Q_{ij},0}, \sigma_{Q_{ij},0}^2, \sigma_{Q_{ij}}^2, \left\{ \eta_{ij,\tau}^S \right\}_{\tau=1}^{T_i-1} \right)$. The corresponding individual-level likelihood is

$$\mathcal{L}_i(\mathbf{y}_i | w_i = k, \mathbf{X}_i; \Theta_i^k) = \prod_{t=1}^{T_i} \prod_{j=1}^J \mathcal{P}(y_{ij,t} = 1 | w_i = k, X_{ij,t})^{y_{ij,t}} \quad (11)$$

where $\mathbf{y}_i = (y_{ij,1}, \dots, y_{ij,T_i})$ and $\mathbf{X}_i = \begin{pmatrix} X_{i1,1} & \dots & X_{iJ,1} \\ \vdots & \ddots & \vdots \\ X_{i1,T_i} & \dots & X_{iJ,T_i} \end{pmatrix}$. Notice that the likelihood function contains two sets of variables that are unobserved by the researcher: the individual-specific process indicator w_i and a series of signal noises $\eta_{ij,\tau}^S$ for $\tau = 1, \dots, T_i - 1$ conditional on $w_i = \infty$. We adopt a Bayesian estimation approach and rely on data augmentation to tackle the issue. These and other related details are discussed next.

3 Data and Estimation

In this section we describe the toothpaste data (containing both scanner panel and survey components) used in our investigation. Particularly, we outline specifics of the stated brand preferences and familiarity information, discuss identification issues and elaborate on our estimation methodology.

3.1 Toothpaste Data

The empirical analysis in this study uses a unique dataset on toothpaste choices and preferences obtained from IRI. The scanner panel data contains 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 dataset. Price is measured as shelf price inclusive of any temporary price discount. In-store display is measured as a scale index ranging from 0 to 1, 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 irrespective of price) 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 by 1 (low) to 7 (high) scale. This stated preference information is valuable because it was collected from the same individuals we have

scanner data on and just before the start of 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 dataset comprises a random sample of 673 households dispersed across the US. 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 4 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 on both liking and familiarity. When compared with Colgate, Crest is priced lower, displayed less frequently, but rated higher on 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 on both liking and familiarity. The medium market share brands - Aquafresh, Mentadent, and Arm & Hammer - generally rank middle in terms of price, display, and survey ratings. There are a couple of noticeable exceptions. Arm & Hammer is the least frequently displayed brand. Mentadent is the highest priced brand and among the most repeatedly purchased brands.

3.2 Familiarity, Preferences and Identification

The composite model proposed in this study is identifiable if each of the component models is identifiable and distinguishable from the others. The Bayesian learning model, one of the component models, is not identifiable in its current form, and we need to impose some additional restrictions to achieve identification. The set of parameters in the Bayesian learning process is $\{Q_{ij}, \sigma_{Q_{ij}}^2, \nu_{Q_{ij},0}, \sigma_{Q_{ij},0}^2\}$ for $\forall i$ and j . In addition, there are a series of unobservable signal noises, $\eta_{ij,\tau}^S$ for $\tau = 1, \dots, T_i - 1$. From equation (7) and (9) 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, $\frac{\sigma_{Q_{ij}}^2}{\sigma_{Q_{ij,0}}^2}$, is identifiable. To resolve this, we set $\sigma_{Q_{ij}}^2 = 1$ for $\forall i$ and j . Consequently, the interpretation of estimated $\sigma_{Q_{ij,0}}^2$ should be relative to $\sigma_{Q_{ij}}^2 = 1$ (e.g., $\widehat{\sigma}_{Q_{ij,0}}^2 = \frac{1}{2}$ means that $\sigma_{Q_{ij,0}}^2$ is a half of $\sigma_{Q_{ij}}^2$). The series of unobservable signal noises will be augmented to the parameter set. Their prior distribution is a product of standard normal densities due to the previous identification restriction.

A remaining question is how one identifies $\nu_{Q_{ij,0}}$ and $\sigma_{Q_{ij,0}}^2$ separately from Q_{ij} . Typical patterns of choices in frequently purchased product categories show that 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 of consumer i . This implies that consumer i 's quality beliefs about the unchosen brands do not evolve over time and consequently, for such a consumer, we cannot distinguish $\nu_{Q_{ij,0}}$ and $\sigma_{Q_{ij,0}}^2$ from Q_{ij} for these brands. In the absence of additional information, we need to impose some restrictions on $\nu_{Q_{ij,0}}$ and $\sigma_{Q_{ij,0}}^2$ to achieve identification.

In our data the survey component provides additional information such as liking and familiarity for each brand. Define $S_{ij} = \{LIK_{ij}, FAM_{ij}\}$ 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 .³ Since 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 as follows. For individuals who are learning,

$$\nu_{Q_{ij,0}} = \bar{\nu}_j + \bar{\phi} \widetilde{LIK}_j, \text{ and} \quad (12)$$

$$\frac{1}{\sigma_{Q_{ij,0}}^2} = \exp(\bar{\kappa} + \bar{\delta} \widetilde{FAM}_{ij}), \quad (13)$$

where $\widetilde{LIK}_{ij} = LIK_{ij} - \frac{1}{N} \sum_{i=1}^N LIK_{ij}$; $\widetilde{FAM}_{ij} = FAM_{ij} - \frac{1}{N} \sum_{i=1}^N FAM_{ij}$; the bar notation over the parameters indicates that they are restricted to be homogeneous among the Bayesian

³In both constructs, 1 implies less and 7 implies more. For example, a 7 on Familiarity would imply that the consumer is very familiar with that particular brand.

learning individuals. Moreover, since liking represents the true mean quality for individuals who are not learning any more, we let the true mean quality be a function of liking, as given by

$$Q_{ij} = \alpha_{ij} + \gamma_i L I K_{ij}. \quad (14)$$

Of course, when survey information is not available we have $\gamma_i = 0$, $\bar{\phi} = 0$ and $\bar{\delta} = 0$. Consequently, $Q_{ij} = \alpha_{ij}$, $\nu_{Q_{ij,0}} = \bar{\nu}_j$ and $\frac{1}{\sigma_{Q_{ij,0}}^2} = \exp(\bar{\kappa})$. In this case, the initial perception bias is pooled across consumers while the initial perception variance is pooled across both consumers and brands. These are standard identification restrictions used in Bayesian learning models applied to scanner panel data (see e.g. Erdem and Keane 1996). The initial market shares of the toothpaste brands before any consumption experience took place 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 purchase behavior and its relationship with quality signals from consumption experience.

The parameter α_{ij} plays a role as the individual level intercept terms at the steady state. Not all brand-specific α_{ij} , as is typical of discrete choice models, are identified so that we set $\alpha_{iJ} = 0$. Moreover, given that α_{ij} and $\bar{\nu}_j$ together serve as the intercept terms at the initial period, not all brand-specific $\bar{\nu}_j$ s are identified and therefore one of them should be locationally fixed (i.e., $\bar{\nu}_J = 0$).⁴ This is the last condition to render the Bayesian learning process fully identified.

Now we redefine the process-specific parameter set as follows: $\Theta_i^0 = \{\alpha_{i1}, \dots, \alpha_{iJ-1}, \gamma_i, \beta_i\}$, $\Theta_i^1 = \{\alpha_{i1}, \dots, \alpha_{iJ-1}, \gamma_i, \beta_i, \lambda_i\}$, and $\Theta_i^\infty = \{\alpha_{i1}, \dots, \alpha_{iJ-1}, \gamma_i, \beta_i, \bar{\nu}_1, \dots, \bar{\nu}_{J-1}, \bar{\phi}, \bar{\kappa}, \bar{\delta}, \eta_{ij,\tau}^S \text{ for } \tau = 1, \dots, T_i - 1\}$. When survey information is not available, γ_i for $\forall i$, $\bar{\phi}$ and $\bar{\delta}$ are set to zero. The process-specific utility specification is then given by

⁴This J needs not be the same as j such that $\alpha_{ij} = 0$.

$$U_{ij,t}^k = \begin{cases} \alpha_{ij} + \gamma_i LIK_{ij} + \beta_i X_{ij,t} + \varepsilon_{ij,t}^U & \text{if } w_i = 0 \\ \alpha_{ij} + \gamma_i LIK_{ij} + \lambda_i y_{ij,t-1} + \beta_i X_{ij,t} + \varepsilon_{ij,t}^U & \text{if } w_i = 1 \\ \alpha_{ij} + \gamma_i LIK_{ij} + \frac{\exp(\bar{\kappa} + \bar{\delta} \widetilde{FAM}_{ij})(\bar{\nu}_j + \bar{\phi} \widetilde{LIK}_{ij}) + \sum_{\tau=1}^{t-1} y_{ij,\tau} \eta_{ij,\tau}^S}{\exp(\bar{\kappa} + \bar{\delta} \widetilde{FAM}_{ij}) + \sum_{\tau=1}^{t-1} y_{ij,\tau}} + \beta_i X_{ij,t} + \varepsilon_{ij,t}^U & \text{if } w_i = \infty \end{cases} \quad (15)$$

The resulting choice probability conditional on the process indicator has the same form as equation (10). Finally, the individual likelihood conditional on the process indicator is expressed as

$$[\mathbf{y}_i | w_i = k, \mathbf{X}_i, \mathbf{S}_i; \Theta_i^k] = \mathcal{L}_i(\mathbf{y}_i | w_i = k, \mathbf{X}_i, \mathbf{S}_i; \Theta_i^k) = \prod_{t=1}^{T_i} \prod_{j=1}^J \left(\frac{\exp(\bar{U}_{ij,t}^k)}{\sum_{q=1}^J \exp(\bar{U}_{iq,t}^k)} \right)^{y_{ij,t}}, \quad (16)$$

where \mathbf{S}_i is individual i 's survey information and the bracket notation $[\cdot]$ is hereafter used for a generic expression of conditional probability distributions.

3.3 MCMC Estimation Scheme

The full parameter space at the individual level Θ_i is represented by the union of order-specific subspaces, each of which can be represented by the product of process indicator $\mathcal{I}(w_i = k)$ and Θ_i^k ,

$$\Theta_i = \bigcup_{k=0,1,\infty} \mathcal{I}(w_i = k) \times \Theta_i^k, \quad (17)$$

where $\Theta_i^0 = \{\alpha_{i1}, \dots, \alpha_{iJ-1}, \gamma_i, \beta_i\}$, $\Theta_i^1 = \{\alpha_{i1}, \dots, \alpha_{iJ-1}, \gamma_i, \beta_i, \bar{\lambda}\}$, and $\Theta_i^\infty = \{\alpha_{i1}, \dots, \alpha_{iJ-1}, \gamma_i, \beta_i, \bar{\nu}_1, \dots, \bar{\nu}_{J-1}, \bar{\phi}, \bar{\kappa}, \bar{\delta}, \{\eta_{ij,\tau}^S\}_{\tau=1}^{t-1}\}$. This type of model specification can be classified as the trans-dimensional model in the literature. The main difficulty of estimation lies in the fact that a MCMC sampler must move both *within* and *between* subspaces Θ_i^k of differing dimensions. A standard Gibbs sampler cannot provide moves between models Θ_i^k without further modification of the setting. It is standard to update $\theta_i^k \in \Theta_i^k$ conditional on $\mathcal{I}(w_i = k)$. However, if one conditions on $\theta_i^k \in \Theta_i^k$, then $\mathcal{I}(w_i = k)$ cannot be updated. To tackle this problem

we adopt the Reversible Jump algorithm proposed by Green (1995). The main idea behind Green's Reversible Jump algorithm is to supplement each of the parameter spaces Θ_i^k with adequate artificial spaces in order to create a bijective mapping between them. The details of how to apply this algorithm to our problem will be presented in Appendix A.

Before proceeding, we complete our hierarchical setup by specifying prior distributions for the parameters. For notational simplicity, we further define $\Psi_i = \{\alpha_{i1}, \dots, \alpha_{iJ-1}, \gamma_i, \beta_i\}$ and $\bar{\Phi} = \{\bar{\nu}_1, \dots, \bar{\nu}_{J-1}, \bar{\phi}, \bar{\kappa}, \bar{\delta}\}$. The former represents a set of the individual level parameters common to all processes, while the latter represents a set of aggregate level parameters specific to the Bayesian learning process. The prior distributions of the model parameters are specified as follows.

1. Process-common individual level parameters $\Psi_i = \{\alpha_{i1}, \dots, \alpha_{iJ-1}, \gamma_i, \beta_i\}$:

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

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

2. Individual level inertia parameter λ_i :

$$[\lambda_i | \bar{\lambda}, \sigma_{\bar{\lambda}}^2] = N(\bar{\lambda}, \sigma_{\bar{\lambda}}^2),$$

$$[\bar{\lambda} | h, H] = N(h, H) \text{ and } [\sigma_{\bar{\lambda}}^2 | g, G] = Inv\chi^2(g, G).$$

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

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

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

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

Hyperparameters $p, P, r, R, h, H, g, G, q_{\bar{\Phi}}$ and $Q_{\bar{\Phi}}$ are appropriately chosen to reflect our diffusive idea of the corresponding prior distributions. These prior distributions, coupled with the likelihood function in (16), specify the posterior distribution conditional on the process indicator. Notice that this conditional posterior distribution is sufficient to design a MCMC sampling procedure for the proposed composite model.

Our sampling procedure starts with an initialization of the MCMC sampler. We draw the starting values of $\bar{\Psi}, V_{\bar{\Psi}}, \bar{\lambda}, \sigma_{\bar{\lambda}}^2, \bar{\Phi}$ and $\{\eta_{ij,\tau}^S\}_{\tau=1}^{t-1}$ from their prior distributions and those of Ψ_i and λ_i from $MVN(\bar{\Psi}, V_{\bar{\Psi}})$ and $N(\bar{\lambda}, \sigma_{\bar{\lambda}}^2)$, respectively. In addition we randomly generate w_i for $\forall i$ so that each individual has an equal chance to follow one of the three candidate processes at the initial iteration. Our sampler then cycles through the following steps with each one performed conditional on current values of all other parameters in the model:

Step 1. Generate proposal moves between and within processes.

Step 2. Accept or reject between-process moves.

Step 3. Update Ψ_i by a M-H sampler.

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

Step 5. Update λ_i by a M-H sampler.

Step 6. Update $\bar{\lambda}$ and $\sigma_{\bar{\lambda}}^2$ by a Gibbs sampler.

Step 7. Update $\bar{\Phi}$ by a M-H sampler.

Step 8. Update $\{\eta_{ij,\tau}^S\}_{\tau=1}^{T_i-1}$ by a M-H sampler.

In steps 1 and 2, the individual-specific process indicator w_i for $\forall i$ is updated using a Reversible Jump algorithm. Conditional on w_i , it is fairly straightforward to conduct all subsequent steps. Sampling procedures in steps 3 and 4 are well-established in the literature since they are the same as those for a standard multinomial logit. Updating the inertia parameter in steps 5 and 6 is also a standard task. Steps 7 and 8 involve updating the parameters specific to the Bayesian learning processes. Narayanan and Manchanda (2006) recently proposed how to conduct a MCMC sampling scheme for the Bayesian learning

model. The sampling procedures in step 7 and 8 are simpler to use because a series of signal noises $\{\eta_{ij,\tau}^S\}_{\tau=1}^{T_i-1}$ are sampled independently and updated simultaneously. Full details on the MCMC sampling scheme are presented in the Appendix A.

Assessing convergence in transdimensional models is not an easy task since parameters have different meanings across the component models (Richardson and Green 1997, Brooks and Giudici 1998). Although all parameters in our case retain the same meaning across all iterations of the sampler, it is still debatable to assess convergence with standard diagnostics such as traceplots and autocorrelation. Our main strategy is to check the performance of our estimation procedure with simulated data and then use such diagnostics after a reasonably large burn-in period. The simulation results show that the MCMC sampler for the proposed composite model converges after several thousand iterations and all parameter estimates recover the true values within sampling error.⁵ Given the validity of our sampler, we collected 50000 draws for our main inference after a burn-in period of 150000 iterations. The standard diagnostics suggested that the above burn-in period was adequate for convergence to be achieved.

4 Results and Empirical Findings

In this section we report our results and discuss empirical findings. These include estimates from the composite model with and without the individual level survey information on familiarity and preferences of the brands. We focus on six areas of interest that pertain to our earlier discussion: (i) parameter estimates and model fit, (ii) brand specific constants and qualities, (iii) sensitivity to marketing mix variables, (iv) magnitude of inertia and learning, (v) process heterogeneity and (vi) individual level insights.

⁵Full details of the simulation results are available from the authors.

4.1 Parameter Estimates and Model Fits

Table 2 provides the fits of the composite model and its special cases while Table 3 provides the parameter estimates of the composite model (with and without survey data). The model fits are measured by the 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 composite model provides a better fit than any of its special cases (in which the households are assumed to follow a single order process). The differences between the composite model log-marginal density and its best fitting component model can be classified as “very strong”.⁶ These results make a strong case for the inclusion of process heterogeneity in traditional brand choice models.

The importance of including survey data is evident from the results. With survey information the respective process order models always fit better. In particular, the fit of the composite model which includes the stated preference information improves from -1543 to -1429. 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 heterogeneous zero-order model, in which the individual level preferences serve only to “shift” the brand specific constants, has a better fit than the composite model which allows process heterogeneity, and hence for inertia and learning, but does not include the survey information (-1501 vs. -1543). Nevertheless, one should not lose sight of the fact that with or without survey the composite model does provide a better fit than the special cases when the population follow just a single order process.

⁶Kass 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.

4.2 Brand Specific Constants and Qualities

The inclusion of the stated preference information has a two-fold impact on the brand specific constants. The pairwise comparisons of these constants (in Table 3) reveal that their mean values across consumers are smaller, and the variances fall dramatically, when stated preferences are included. These reduction in the means and in particular in the heterogeneity of the brand specific constants are another indication that the stated preferences provide valuable information on the variation in the true mean qualities.

In Figure 1 we plot the individual posterior means of the true mean qualities, which are defined as in equation (14). The estimated true mean qualities with survey information are for all brands more dispersed. This suggests that most of the individual level variations contained in the stated preference information is not captured by the brand specific constants. Some of that variation, in the survey-less case, is carried over to other constructs in the model which are correlated with the unknown individual preferences, such as past purchase behavior.

4.3 Sensitivity to Marketing Mix Variables

The mean effect and heterogeneity in the sensitivity of marketing mix variables, reported in Table 2, is different when stated preference information is accounted for. Without survey information consumers are (on average) thought to be more price sensitive and more dispersed than they actually are. This happens partially because the absence of preference information forces price to account for more than its true effect. Moreover, it seems that the consumer whose rank preference ordering is known is less willing to switch away from his most preferred brand in response to a competitive promotion, than the “average” consumer. We note, however, that while the aggregate marketing mix effects seem reduced there may be individual cases where the effects move in the opposite direction (larger effects with survey data). Finally, it should be noticed that in comparing sensitivities to price changes across models they depend not only on the price coefficients but also on the brand specific constants and true mean qualities which they need to “overcome”. Since the latter vary across brands

so will the actual sensitivities. Display effects, on average, have a larger mean but similar to price exhibit somewhat lower variances when survey data is included. Figure 2 depicts the individual posterior means of price and display. We will return to these issues in later sections dealing with individual level insights and counterfactual experiments.

4.4 Magnitude of Inertia and Learning

The coefficient for the lagged choice dummy is reported in Table 2 and its individual posterior mean is provided in Figure 3. The individual posterior mean for inertia is larger and more dispersed without the survey information. That is consistent with the notion that without the individual level survey information the last purchase being individual specific is indicative of both inertia and preference.

In terms of learning the brand specific individual posterior means of the initial perception biases are shown in Figure 4. Without survey information, initial perception biases for the brands are more negative and less heterogeneous, indicating the larger amount of learning. The extent of learning is determined not only by the initial perception biases but also by variances. Presented in Figure 5 is the joint impact of these parameters on learning. The predicted average learning during 10 consecutive purchases of each brand is lower and slower when the preference information is accounted for. Furthermore, this finding is more pronounced for the large share brands such as Colgate and Crest. While consumers actually stopped learning about Crest even from the very beginning, their choices of Crest, in the absence of survey data, are at least partially attributed to learning. We provide more detailed discussion of individual level learning in a later section.

Our findings related to both the brand specific constants and qualities and the magnitude of inertia and learning indicate that without the survey data inertia and learning (which are based on an individual's past purchase behavior) act, in part, as proxies for individual deviations from the average preferences, measured via the brand specific constants. Or in other words, inertia and learning are at least partially serving as the basis for the researcher learning about the consumer's unknown preferences. While these effects underline the impor-

tance of process heterogeneity and preference information at the aggregate level, the impact of these constructs in explaining individual behavior is even more striking. We turn to that discussion next.

4.5 Process Heterogeneity

The bottom few rows of Table 3 provide the posterior means for each process indicator. Broadly speaking, these values represent the tendency of the “average” consumer to be of a particular order type. Without the preference and familiarity information Bayesian Learning (0.3742) seems to best describe the average consumer by beating out both the zero-order (0.3120) and inertia (0.3137) process. However, once the survey information has been added there is a marked shift in the process indicators. The zero-order process becomes much more likely (0.3949) mostly at the expense of the Bayesian Learning process (down to 0.3205) and the first-order process as well (down to 0.2846). The first-order process remains relatively stable with a mean of 0.3137 without the survey data and 0.2846 with survey data. This suggests that first-order behavior might be more cleanly estimated even in the absence of preference and familiarity information.

A key feature of our estimation approach is the ability to uncover individual level parameters. In particular we are able to recover individual level marketing mix sensitivities (price and display), true quality estimates and learning parameters. Of particular interest, are the process indicators which measure the proclivity of a given consumer towards zero-order, first-order or Bayesian learning type behavior. The concordance matrix depicted in Table 4 sketches the classification of consumers’ choice process in both the survey and without survey data cases. The classification in this table is based on assigning consumer’s to that order that had the highest mean posterior probability.

There are several noteworthy elements in Table 4: First, in the absence of preference information 32.5% of the population is classified as zero-order, 37.0% as first-order, and 30.5% as Bayesian learners. With survey data the proportion of first-order individuals diminishes slightly to 34.8% but a major shift occurs in the zero-order and Bayesian learning mixture.

The proportion of zero-order jumps to 41.5% while that of Bayesian learners falls to 24.3%. Second, about 30.0% (off diagonal cells) of consumers were erroneously classified into the wrong choice process bin in the absence of survey data. The majority of misclassifications (66.4%) are in the lower triangle of the matrix suggesting that in the absence of preference and familiarity information the order of the choice process is likely to be overestimated. The largest proportion (40.4%) of the misclassifications occur on account of the fact that the absence of preference information confounds zero-order behavior and Bayesian learning behavior. Or in other words, a large fraction of the individuals who are classified without survey data as Bayesian learners are actually zero-order individuals about whom the researcher is learning. Finally, the first-order process seems to be cleanly identified even in the absence of preference data (i.e., 88% are correctly classified).

4.6 Individual Level Insights

The composite model brings together various sources of heterogeneity: (a) Heterogeneous learning and (b) Process and Parameter heterogeneity. We take a deeper look at each, and discuss their implications, in turn.

4.6.1 Individual Learning Behavior

The earlier discussion outlined the distribution of process heterogeneity in the sample. We can, however, go deeper and examine individual level learning patterns. Figure 6 plots the learning behavior of all individuals classified as being Bayesian learners in the with and without survey cases. There are several noteworthy differences between the two. First, the graphs attest to our earlier discussion that the number of individuals classified as learners is much larger in the absence of survey data on preferences and familiarity. Second, the plots show that the nature of learning across the two data regimes is also very different. While there seems to be substantial learning without the survey data, there is much less so when preferences are included in the model. This is consistent with our aggregate learning results. What is striking though is that there is much larger *heterogeneity* in learning behavior when

preferences are included. This is evidenced by the spread and overlap of the learning curves. Finally, a key takeaway from this plot is that even though individuals are classified as learners in the with-survey case, for most consumers there is very little distinction between learning and zero-order behavior. This has important implications for managers which we will discuss in the sequel.

4.6.2 Process and Parameter Heterogeneity

Table 5 depicts three households with different choice patterns and preferences facing varied marketing mix environments. For each consumer the table also presents the estimated (mean) posterior probability of the process indicators and the price coefficient. Household #297 is the motivating example (named Jane) we introduced in the introduction to this paper. Simply examining the scanner part of the data, it should come as no surprise that Jane is thought of as a first order consumer. She buys Aquafresh (AF) repeatedly and even when the price goes above the mean price. It is only when the price of AF is significantly above the mean level that she switches over to Arm and Hammer (AH). Since the choice pattern supports a first order behavior she is classified as so with a high probability (about 83%). A quick examination of the survey data information on liking and familiarity tells a very different story. Aquafresh is Jane's most preferred brand (by far) and also the one she is most familiar with. Given this information, it is obvious that Jane is buying Aquafresh not because of some inertial component but simply because she likes the brand! In other words, she is a zero-order type. Since the preferences explain a large proportion of the choice patterns it also explains why the price coefficient with the survey data is now less negative.

Household #27 presents a more complex scenario. Without the survey data the choice patterns exhibit higher order behavior (e.g. Colgate (CG) is purchased in the beginning and then again after a few occasions, the same for Aquafresh (AF) and Crest (CR).) Given these patterns the model ascribes a Bayesian learning tag to this household. Since the brand switching also seems to be very related to prices the price coefficient is relatively large. With survey data we have more information about preferences and again a different picture

emerges. The choice of Colgate (CG) is mostly explained by the preference and familiarity, while the choice of the other brands is explained by a combination of price and preferences (PS, AF and CR). Given this information, the process posterior probabilities shift around to significantly favor zero order behavior. Further, the price coefficient is significantly reduced (towards zero) since prices are not the only factor explaining the brand switches. Both the above examples highlight the overestimation of the choice process in the absence of survey information.

As a final example we focus on Household # 119 where both with and without preference/familiarity information the order of the choice process is estimated to be zero order. It is straightforward to see why: The scanner data clearly shows that the household switches away from Mentadent (MT) only when MT is priced high and in such case the household consistently switches to Crest (CR). The survey data, in this case, does not add a large amount of extra information (although knowing the preferences allow us to more precisely ascertain the choice process). An interesting side issue here is that the price coefficient in the with survey case is more negative than the without survey case. This happens because, on the margin, the brand switches are no longer being explained by preferences but rather by prices.

Our individual level analysis uncovered many more examples which offer insights similar to those presented in these examples. For the sake of brevity we have limited ourselves to three cases.

5 Managerial Implications

In the previous section it became evident that when survey data is included, consumers are found to be less sensitive to marketing mix activities as well as their own past purchase behaviors. In addition the individual posterior true mean qualities of the brands, which a change in a marketing mix activity would need “overcome”, are based on Figure 1 much more dispersed when survey data is available. In this section, in order to assess the economic and managerial implications of our empirical findings, we carried out three counterfactual

experiments, namely, free-sampling, coupon, and in-store display. Moreover, to ascertain the value of survey information, each experiment is conducted twice: with and without survey information.

5.1 Counterfactual Experiments

In the free-sampling experiment, a brand is assumed to distribute a free-sample to each consumer just before the initial observation period. The information content of a free-sample is assumed to be equivalent to that of a regular product. In the coupon experiment, a brand is assumed to distribute a 50-cent coupon to each consumer just before the initial observation period. We assume that all consumers exercise the coupon at their first purchase occasion. Lastly, in the in-store display experiment, a brand is assumed to engage in display activity at full intensity during the first twelve weeks. In all cases, no competitive reaction is allowed.

In these sales promotion experiments, we revisit each consumer’s purchase occasions and simulate “baseline” and “post-promotion” brand choices by using the actual marketing mix information and the simulated errors. This procedure is conducted for all brands one by one. For each brand the incremental own market shares (i.e., post-promotion market share minus baseline market share) and the corresponding revenue gains are calculated.

Presented in Table 6 are results of the aforementioned three experiments. In almost all cases, the effects of sales promotion tend to be overly optimistic in the absence of survey information. That is, the predicted incremental market shares and revenue are highly inflated (and on the surface unreasonable) when the stated preference information collected through survey are not available. These prediction gaps are more salient for the large share brands such as Colgate and Crest. For instance, without survey information, a free-sample of Crest to all consumers increases its market share by 6.55% and its revenue by \$420.74, while with survey information the increase is only to 1.37% for market share and \$89.79 for revenue. A 50-cent coupon of Crest distributed to and exercised by all consumers, without survey information, is predicted to induce a 6.72% increase of market share and a \$338.12

revenue gain. With survey information the corresponding numbers are 3.76% and \$156.37, respectively. The in-store display experiment for Crest also provides evidence of similar kind.

These counterfactual experiments illustrate that our empirical findings based on the parameter estimates (e.g., overestimation of state dependence and bias in sensitivity to marketing mix variables) are also economically significant. Consequently, any sales promotion policy based solely on scanner panel data is likely to be flawed and based on the results reported in Table 6 lead to unreasonable managerial expectations.

5.2 Takeaways

The results in this paper have some key takeaways for practicing managers. First, accounting for process heterogeneity is important. Naively assuming that all consumers have the same process order creates significant biases in the parameter estimates. Further, these biases translate into significant dollar differences as the previous section shows. Second, having data on individual preferences alleviates significantly the problems that crop up due to the confounding of preference heterogeneity and state-dependence. Once again, without such data very different process and parameter effects are uncovered. For example, not only are more consumers likely to be perceived as learning but also the degree to which they are learning will be greatly overstated. This has important implications for dynamic pricing and promotion strategies. Our conjecture is that the investment costs incurred in the collection of such survey information will be more than offset by the benefits accruing from more precise estimates.

6 Summary and Conclusion

Consumers in choosing brands within a product category act intelligently. They use their existing preferences, which are based on multi-attribute utilities, and update those based on their own consumption experiences. Researchers who study consumers, in the context of frequently purchased product categories, have expressed explicitly or implicitly wide ranging

opinions as to how consumers made their choices. Some believe that given the frequency of purchase consumers have all the relevant attribute information and have converged to fixed purchase probabilities. The reasons that those consumers switch altogether is either due to inherent randomness or that a mixture of the chosen brands provides the right attribute mix. Others believe that consumers, in addition to attribute information, are impacted by inertia/last purchase reinforcement. That is, following a purchase of a brand its utility goes up but will come down if it is not purchased after that again; another brand chosen at that time will get the temporary spike in utility. Yet others believe that consumers continuously update their preferences based on consumption experiences in a very systematic and sophisticated manner (Bayesian updating). A possibility totally ignored in the scanner based logit analysis literature is that all of the above scholars may be right. Within the same product category some consumers may follow each of the above processes. The exact mixture of consumer types will depend on the stability of the product class in terms of the introduction of new brands, brand repositioning, the influx of new consumers, and the like.

In this study we investigated such a mixture of consumer types through a composite model which allows for both process heterogeneity and within process parameter heterogeneity, and indeed found that all process types exist side by side. While it might be argued that we did so only for one product category similar results in other product categories were identified in a couple of much earlier articles (reviewed in the introduction) which used diary panel data. In several ways this study provides more powerful evidence for the existence of these process orders. Unlike the earlier studies the current one (just like other scanner data studies) has information on changes in marketing mix variables, such as price and display, in addition, it is unique in that it also contains information on individuals' brand familiarities and preferences. This latter information was found in our analysis to be invaluable. An analysis without it identified too many individuals as learners, too few as zero-order, and on average, consumers to be overly sensitive to marketing mix variables.

Based on our analysis without the survey information the extent of state dependence, the mean probability of the inertial and learning indicators and the amount of learning is

spuriously overestimated. That is, part of what is identified as consumer state dependence is actually the researcher learning about the unknown consumer preferences. Nevertheless, it needs to be stressed that we did find that consumer inertia and learning effects exist and are substantial even when preferences are known in an established product category such as toothpaste. Moreover, through counterfactual experiments we were able to show, for example, that the magnitude of state dependence is sufficiently large to make a price promotion profitable. While we demonstrated this for a case in which there were no competitive reactions to a brand's price promotion we have reason to believe that given the existence of state dependence, as shown analytically by Freimer and Horsky (2003), competitive promotions would also be optimal. Clearly they are predicted by the brand managers in this category.

One of our overall recommendations, with implications for both researchers and managers, is to use, despite its complexity and the involved estimation methodology, a composite model, such as forwarded in this study, and thus to allow for the existence of consumers process order types. A priori specifying only a single choice process for all consumers severely underestimates the underlying consumer heterogeneity while also biasing other key effects. Moreover, we highly recommend that the choice data be augmented with familiarity and preference information. In order to correctly assess the size (and composition) of the segments and the sensitivity of the population to marketing mix variables, such survey based information seems crucial.

References

- Ackerberg, D. (2003), "Advertising, Learning, and Consumer Choice in Experienced Good Markets: An Empirical Examination", *International Economic Review* 44, 1007-1040.
- Aliwadi, K. L., K. Gedenk, and S. Neslin (1999), "Heterogeneity and Purchase Event Feedback in Choice Models: An Empirical Analysis with Implications for Model Building", *International Journal of Research in Marketing* 16, 177-198.
- Allenby, G. M. and P. J. Lenk (1995), "Reassessing Brand Loyalty, Price Sensitivity and Merchandising Effects on Consumer Brand Choice", *Journal of Business and Economic Statistics* 13, 281-289.
- Blattberg, R. C. and S. K. Sen (1976), "Market Segments and Stochastic Brand Choice Models", *Journal of Marketing Research* 13, 34-45.
- Bass, F. M. (1974), "The Theory of Stochastic Preference and Brand Switching", *Journal of Marketing Research* 11, 1-20.
- Bass, F. M., M. Givon, M. U. Kalwani, D. Reibstein, and G. P. Wright (1984), "An Investigation into the Order of the Brand Choice Process", *Marketing Science* 3, 267-287.
- Bass, F. M. and J. Wind (1995), "Introduction to the Special Issue: Empirical Generalizations in Marketing", *Marketing Science* 14, G1-5.
- Brooks, S. P. and P. Giudici (1998), *Convergence Assessment for Reversible Jump MCMC Simulations*, in Bayesian Statistics, Vol. 6., Oxford University Press, Oxford.
- Ching, A. (2000), "Dynamic Equilibrium in the US Prescription Drug Market After Patent Expiration", Ph.D. dissertation, University of Minnesota.
- Chintagunta, P. K., D. C. Jain, and N. J. Vilcassim (1991), "Investigating Heterogeneity in Brand Preferences in Logit Models for Panel Data", *Journal of Marketing Research* 28, 417-428.
- Eckstein, Z., D. Horsky, and Y. Raban (1988), "An Empirical Dynamic Model of Optimal Brand Choice", Working Paper No. 88, Tel Aviv University.
- Erdem, T. and M. P. Keane (1996), "Decision-making Under Uncertainty: Capturing Dynamic Brand Choice Processes in Turbulent Consumer Goods Markets", *Marketing Science* 15, 1-20.
- Frank, R. E. (1962), "Brand Choice as a Probability Process", *Journal of Business* 35, 43-56.
- Freimer, M. and D. Horsky (2003), "Try It You Will Like It - Does Consumer Learning Lead to Competitive Price Promotion", Working Paper, University of Rochester.

- Gilbride, T. J., G. M. Allenby, and J. Brazell (2006), "Models of Heterogeneous Variable Selection", *Journal of Marketing Research* 43, 420-430.
- Givon, M. and D. Horsky (1979), "Application of a Composite Stochastic Model of Brand Choice", *Journal of Marketing Research* 16, 258-267.
- Givon, M. and D. Horsky (1985), "Homogeneity and Heterogeneity in Stochastic Models of Brand Choice Behavior", *International Journal of Research in Marketing* 2, 263-272.
- Gönül, F. and K. Srinivasan (1993), "Modeling Multiple Sources of Heterogeneity in Multinomial Logit Models: Methodological and Managerial Issues", *Marketing Science* 12, 213-229.
- Green, P. J. (1995), "Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination", *Biometrika* 82, 711-732.
- Guadagni, P. M. and J. D. Little (1983), "A Logit Model of Brand Choice Calibrated on Scanner Data", *Marketing Science* 2, 203-238.
- Hardie, B., E. Johnson, and P. Fader (1993), "Modelling Loss Aversion and Reference Dependence Effects on Brand Choice", *Marketing Science* 12, 378-394.
- Heckman, J. J. (1991), "Identifying the Hand of the Past: Distinguishing State Dependence from Heterogeneity", *American Economic Review* 81, 75-79.
- Herniter, J. D. (1973), "An Entropy Model of Brand Purchase Behavior", *Journal of Marketing Research* 10, 361-375.
- Horsky, D., S. Misra, and P. Nelson (2006), "Observed and Unobserved Heterogeneity in Brand-Choice Models", *Marketing Science* 25, 322-335.
- Jones, J. M. (1973), "A Composite Heterogeneous Model for Brand Choice Behavior", *Management Science* 19, 499-509.
- Jeuland, A. P. (1979), "Brand Choice Inertia as One Aspect of the Notion of Brand Loyalty", *Management Science* 25, 671-682.
- Kass, R. E. and A. E. Raftery (1995), "Bayes Factors", *Journal of the American Statistical Association* 90, 773-795.
- Kamakura, W. A. and G. J. Russell (1989), "A Probabilistic Choice Model for Market Segmentation and Elasticity Structure", *Journal of Marketing Research* 26, 379-390.
- Keane, M. P. (1997), "Modelling Heterogeneity and State Dependence in Consumer Choice Behavior", *Journal of Business and Economic Statistics* 15, 310-327.
- Kuehn, A. A. (1962), "Consumer Brand Choice - A Learning Process?", *Journal of Advertising Research* 2, 10-17.

- Massy, W. F., D. B. Montgomery, and D. G. Morrison (1970), *Stochastic Models of Buying Behavior*, MIT Press.
- Mehta, N., S. Rajiv, and K. Srinivasan (2003), "Price Uncertainty and Consumer Search: A Structural Model of Consideration Set Formation", *Marketing Science* 22, 58-84.
- Morrison, D. G. (1966), "Testing Brand-Switching Models", *Journal of Marketing Research* 3, 401-409.
- Narayanan, S. and P. Manchanda (2005), "Heterogeneous Learning and the Targeting of Marketing Communications for New Products", Working Paper, University of Chicago.
- Newton, M. A. and A. E. Raftery (1994), "Approximate Bayesian Inference with the Weighted Likelihood Bootstrap", *Journal of the Royal Statistical Society, Series B* 56, 3-48.
- Richardson, S. and P. J. Green (1997), "On Bayesian Analysis of Mixtures with an Unknown Number of Components", *Journal of the Royal Statistical Society, Series B* 59, 731-758.
- Rossi, P. E., R. E. McCulloch and G. Allenby (1996) "The Value of Purchase History Data in Target Marketing" *Marketing Science*, 15, 321-340.
- Roy, R., P. K. Chintagunta, and S. Haldar (1996), "A Framework for Investigating Habits, "the hand of the Past", and Heterogeneity in Dynamic Brand Choice", *Marketing Science* 15, 280-299.
- Seetharaman, P. B. and P. K. Chintagunta (1998), "A Model of Inertia and Variety-seeking with Marketing Variables", *International Journal of Research in Marketing* 15, 1-17.
- Seetharaman, P. B., A. K. Ainslie, and P. K. Chintagunta (1999), "Investigating Household State Dependence Effects across Categories", *Journal of Marketing Research* 36, 488-500.
- Seetharaman, P. B. (2003a), "Probabilistic versus Random-utility Models of State Dependence: An Empirical Comparison", *International Journal of Research in Marketing* 20, 87-96.
- Seetharaman, P. B. (2003b), "Modeling Multiple Sources of State Dependence in Random Utility Models: A Distributed Lag Approach", *Marketing Science* 23, 263-271.
- Uncles, M., A. Ehrenberg, and K. Hammond (1995), "Patterns of Buyer Behavior: Regularities, Models, and Extensions", *Marketing Science* 14, G71-78.

Appendix A (MCMC Implementation Details)

Our sampler consists of a sequence of Gibbs, Metropolis-Hastings, and Reversible Jump steps. For fitting the composite model of process heterogeneity, we conduct a Reversible Jump step to sample across component processes (e.g., accept/reject moves between any pair of three component processes) and a Metropolis/Gibbs step within each component process. We here illustrate the MCMC sampling procedure outlined in the estimation section. Following is the details of each of eight steps employed to estimate the proposed model in this study.

Step 1: Generate proposal moves between and within processes.

For each individual i , we draw a proposal value of the process indicator w'_i according to the move transition probability that $\Pr(w'_i = q|w_i = p) = \frac{1}{3}$ for $p = 0, 1, \infty$ and $q = 0, 1, \infty$. For example, it is equally likely that those whose current value of w_i is equal to 0 are proposed to stay within the zero-order process ($w'_i = 0$) or switch to either the inertia process ($w'_i = 1$) or the Bayesian learning process ($w'_i = \infty$). The same is true for those whose current value of w_i is equal to 1 or ∞ .

Step 2: Accept or reject between-process moves.

Moves between models involve changing the number of parameters and thus adding new parameters or removing older ones. The acceptance ratio for between-process moves is generally defined as $\min\{1, r\}$ where $r = \text{likelihood ratio} \times \text{prior ratio} \times \text{proposal ratio} \times \text{Jacobian}$.

There are three possibilities of between-process moves: between the zero-order and inertia processes, between the zero-order and Bayesian learning processes, and between the inertia and Bayesian learning processes.

1. Between the zero-order and inertia processes

To jump from the zero-order process to the inertia process (i.e., $\Psi_i \rightarrow \{\Psi'_i, \lambda'_i\}$), we have to draw the auxiliary random variable u for each individual from a proposal density denoted by $J^\lambda(u)$. Then we define the value of the inertia parameter by setting $\lambda'_i = u$ and leaving Ψ_i as they are in the current iteration (i.e., $\Psi'_i = \Psi_i$). According to the above template proposed by Green (1995), the acceptance probability of this move is

$$\min\left(1, \frac{[\mathbf{y}_i|w_i = 1, \mathbf{X}_i, \mathbf{S}_i; \Psi_i, \lambda'_i][\lambda'_i|\bar{\lambda}, \sigma_{\bar{\lambda}}^2]}{[\mathbf{y}_i|w_i = 0, \mathbf{X}_i, \mathbf{S}_i; \Psi_i]J^\lambda(\lambda'_i)}\right).$$

Notice that the prior distribution of Ψ_i is cancelled out and the Jacobian of the transformation is equal to one. To jump from the inertia process to the zero-order process (i.e., $\{\Psi_i, \lambda_i\} \rightarrow \Psi'_i$), we merely set $\lambda_i = 0$ and retain the current values of Ψ_i (i.e., $\Psi'_i = \Psi_i$). The acceptance probability of this reversal move is the reciprocal of the above.

2. Between the zero-order and Bayesian learning processes

To jump from the zero-order process to the Bayesian learning process (i.e., $\Psi_i \rightarrow \{\Psi'_i, \bar{\Phi}', \{\eta_{ij,\tau}^{S'}\}_{\tau=1}^{T_i-1}\}$), we have to draw the auxiliary random variables for $\bar{\Phi}'$ and $\{\eta_{ij,\tau}^{S'}\}_{\tau=1}^{T_i-1}$. We propose a vector of the candidate values u for $\bar{\Phi}$ from a jumping distribuion denoted by $J^\Phi(u)$ and use the prior distribution of $\eta_{ij,\tau}^S$ as a proposal density for $\eta_{ij,\tau}^{S'}$. Then we define the candiate values of the learning parameters by setting $\bar{\Phi}' = u$ and leaving Ψ_i as they are in the current iteration (i.e., $\Psi'_i = \Psi_i$). The acceptance probability of this move is

$$\min(1, \frac{[\mathbf{y}_i|w_i = \infty, \mathbf{X}_i, \mathbf{S}_i; \Psi_i, \bar{\Phi}', \{\eta_{ij,\tau}^{S'}\}_{\tau=1}^{T_i-1}][\bar{\Phi}'|q_{\bar{\Phi}}, Q_{\bar{\Phi}}]}{[\mathbf{y}_i|w_i = 0, \mathbf{X}_i, \mathbf{S}_i; \Psi_i]J^\Phi(\bar{\Phi}')}).$$

Notice again that the prior distribution of Ψ_i is cancelled out and the Jacobian of the transformation is equal to one. In addition, the prior density of $\{\eta_{ij,\tau}^S\}_{\tau=1}^{T_i-1}$ in the numerator is cancelled out by their proposal density in the denominator. To jump from the Bayesian learning process to the zero-order process, $\{\Psi_i, \bar{\Phi}, \{\eta_{ij,\tau}^S\}_{\tau=1}^{T_i-1}\} \rightarrow \Psi'_i$, we merely set $\bar{\Phi} = 0$ and $\{\eta_{ij,\tau}^S = 0\}_{\tau=1}^{T_i-1}$ and retain the current values of Ψ_i (i.e., $\Psi'_i = \Psi_i$). The acceptance probability of this reversal move is again the reciprocal of the above.

3. Between the inertia and Bayesian learning processes

The construction of this move is a mixture of the previous two. To jump from the inertia process to the Bayesian learning process (i.e., $\{\Psi_i, \lambda_i\} \rightarrow \{\Psi'_i, \bar{\Phi}', \{\eta_{ij,\tau}^{S'}\}_{\tau=1}^{T_i-1}\}$), we retain the current values of Ψ_i , propose the candiate values for $\bar{\Phi}$ and $\{\eta_{ij,\tau}^S\}_{\tau=1}^{T_i-1}$, and set $\lambda_i = 0$. The acceptance probability of this move is

$$\min(1, \frac{[\mathbf{y}_i|w_i = \infty, \mathbf{X}_i, \mathbf{S}_i; \Psi_i, \bar{\Phi}', \{\eta_{ij,\tau}^{S'}\}_{\tau=1}^{T_i-1}][\bar{\Phi}'|q_{\bar{\Phi}}, Q_{\bar{\Phi}}]J^\lambda(\lambda_i)}{[\mathbf{y}_i|w_i = 1, \mathbf{X}_i, \mathbf{S}_i; \Psi_i, \lambda_i][\lambda_i|\bar{\lambda}, \sigma_\lambda^2]J^\Phi(\bar{\Phi}')}).$$

The reversal move is constructed in a similar fashion to the previous cases. Its acceptance probabiity is equal to the reciprocal of the above.

Efficient constructon of the jumping distributions in the reverible jump context is a challenging task. In this study we choose $J^\lambda(u)$ and $J^\Phi(u)$ based on a pilot analysis of the corresponding component models. For example, we construct $J^\lambda(u)$ using a normal distribution whose mean and variance match the posterior mean and variance from a single-order inertia model. $J^\Phi(u)$ is constructed similarly.

Step 3: Update $\Psi_i = \{\alpha_{i1}, \dots, \alpha_{iJ-1}, \gamma_i, \beta_i\}$ by a M-H sampler.

The full conditional distribution of Ψ_i is

$$[\Psi_i|rest] \propto \left(\begin{array}{l} [\mathbf{y}_i|w_i = 0, \mathbf{X}_i, \mathbf{S}_i; \Psi_i]^{\mathcal{I}(w_i=0)} \\ \times [\mathbf{y}_i|w_i = 1, \mathbf{X}_i, \mathbf{S}_i; \Psi_i, \lambda_i]^{\mathcal{I}(w_i=1)} \\ \times [\mathbf{y}_i|w_i = \infty, \mathbf{X}_i, \mathbf{S}_i; \Psi_i, \bar{\Phi}, \{\eta_{ij,\tau}^S\}_{\tau=1}^{T_i-1}]^{\mathcal{I}(w_i=\infty)} \end{array} \right) [\Psi_i|\bar{\Psi}, V_{\bar{\Psi}}],$$

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 $\min(1, \frac{[\Psi'_i|rest]}{[\Psi_i|rest]})$. This step is conducted on an individual basis.

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

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

$$\begin{aligned} [\bar{\Psi}|rest] &= N\left(\frac{V_{\bar{\Psi}}^{-1} \sum_{i=1}^I \Psi_i + P^{-1}p}{V_{\bar{\Psi}}^{-1}I + P^{-1}}, (V_{\bar{\Psi}}^{-1}I + P^{-1})^{-1}\right) \text{ and} \\ [V_{\bar{\Psi}}|rest] &= IW(r + I, (R + \sum_{i=1}^I ((\Psi_i - \bar{\Psi})(\Psi_i - \bar{\Psi}'))^{-1}), \end{aligned}$$

from which it is straightforward to sample.

Step 5: Update λ_i by a M-H sampler.

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

$$[\lambda_i|rest] \propto \left(\prod_{i=1}^I [\mathbf{y}_i | w_i = 1, \mathbf{X}_i, \mathbf{S}_i; \Psi_i, \bar{\lambda}]^{\mathcal{I}(w_i=1)} \right) [\lambda_i | \bar{\lambda}, \sigma_{\bar{\lambda}}^2],$$

and we generate a proposal value λ'_i using a symmetric random walk M-H algorithm. The acceptance probability of λ'_i is $\min(1, \frac{[\lambda'_i|rest]}{[\lambda_i|rest]})$. Notice that this step is conducted only for those whose $w_i = 1$.

Step 6: Update $\bar{\lambda}$ and $\sigma_{\bar{\lambda}}^2$ by a Gibbs sampler.

Due to the conjugate prior specification for $\bar{\lambda}$ and $\sigma_{\bar{\lambda}}^2$, their full conditional distributions are

$$\begin{aligned} [\bar{\lambda}|rest] &= N\left(\frac{\sigma_{\bar{\lambda}}^{-2} \sum_{i=1}^I \lambda_i \mathcal{I}(w_i = 1) + H^{-1}h}{\sigma_{\bar{\lambda}}^{-2} \sum_{i=1}^I \mathcal{I}(w_i = 1) + H^{-1}}, (\sigma_{\bar{\lambda}}^{-2} \sum_{i=1}^I \mathcal{I}(w_i = 1) + H^{-1})^{-1}\right) \text{ and} \\ [\sigma_{\bar{\lambda}}^2|rest] &= Inv - \chi^2\left(g + \sum_{i=1}^I \mathcal{I}(w_i = 1), \frac{gG + \sum_{i=1}^I (\lambda_i - \bar{\lambda})^2 \mathcal{I}(w_i = 1)}{g + \sum_{i=1}^I \mathcal{I}(w_i = 1)}\right), \end{aligned}$$

from which it is straightforward to sample. Notice that only those whose $w_i = 1$ contribute to the full conditional distributions of $\bar{\lambda}$ and $\sigma_{\bar{\lambda}}^2$.

Step 7: Update $\bar{\Phi}$ by a M-H sampler.

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

$$[\bar{\Phi}|rest] \propto \prod_{i=1}^I [\mathbf{y}_i | w_i = \infty, \mathbf{X}_i, \mathbf{S}_i; \Psi_i, \bar{\Phi}, \{\eta_{ij,\tau}^S\}_{\tau=1}^{T_i-1}]^{\mathcal{I}(w_i=\infty)} [\bar{\Phi} | q_{\bar{\Phi}}, Q_{\bar{\Phi}}],$$

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 $\min(1, \frac{[\bar{\Phi}'|rest]}{[\bar{\Phi}|rest]})$. Notice that only those whose $w_i = \infty$ contribute to the full conditional distribution of $\bar{\Phi}$.

Step 8: Update $\{\eta_{ij,\tau}^S\}_{\tau=1}^{T_i-1}$ by a M-H sampler.

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

$$[\{\eta_{ij,\tau}^S\}_{\tau=1}^{T_i-1}|rest] \propto [\mathbf{y}_i|w_i = \infty, \mathbf{X}_i, \mathbf{S}_i; \Psi_i, \bar{\Phi}, \{\eta_{ij,\tau}^S\}_{\tau=1}^{T_i-1}] \prod_{\tau=1}^{T_i-1} [\eta_{ij,\tau}^S | \mu_\eta, \sigma_\eta^2],$$

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 $\min(1, \frac{[\mathbf{y}_i|w_i=\infty, \mathbf{X}_i, \mathbf{S}_i; \Psi_i, \bar{\Phi}, \{\eta_{ij,\tau}^{S'}\}_{\tau=1}^{T_i-1}]}{[\mathbf{y}_i|w_i=\infty, \mathbf{X}_i, \mathbf{S}_i; \Psi_i, \bar{\Phi}, \{\eta_{ij,\tau}^S\}_{\tau=1}^{T_i-1}]})$. This step is conducted on an individual basis and only for those whose $w_i = \infty$.

Table 1: Descriptive Statistics for Toothpaste Data

Variable		Brand						
		Arm & Hammer	Aim	Aquafresh	Colgate	Crest	Mentadent	Pepsodent
Market share		0.0704	0.0260	0.1523	0.3179	0.3123	0.1052	0.0160
Repeated Purchase Probability		0.4650	0.3750	0.5138	0.5725	0.5759	0.5764	0.2813
Price	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
Display	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
Liking	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
Familiarity	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

Table 2: Log Marginal Densities of the Estimated Models

Model	Data	
	Without Survey	With Survey
Component Models		
<i>Zero-order</i>	-1692	-1501
<i>Inertia (First-order)</i>	-1643	-1462
<i>Bayesian Learning</i>	-1616	-1451
Proposed Composite Models	-1543	-1429

Table 3 : Parameter Estimates of the Proposed Composite Model

Parameter	Standard Scanner Data (Choice+Marketing mix)		Survey Augmented Data (Choice+Marketing mix+Survey)	
	Posterior Mean	Heterogeneity [†]	Posterior Mean	Heterogeneity [†]
<u>True Mean Quality</u>				
<i>Arm & Hammer</i>	8.9432 (8.3311, 9.4417)	1.3518 (0.9396, 1.6869)	7.2185 (6.6479, 7.8872)	0.7092 (0.3965, 1.0853)
<i>Aim</i>	1.6618 (1.1593, 2.1627)	1.1800 (0.7383, 1.6892)	0.4045 (-0.3170, 0.9073)	0.6960 (0.4378, 0.9325)
<i>Aquafresh</i>	8.4634 (7.9795, 8.9097)	1.1925 (0.8022, 1.5254)	6.2913 (5.8311, 6.9041)	0.8495 (0.4427, 1.2707)
<i>Colgate</i>	10.5835 (10.1627, 11.0966)	1.6969 (1.2185, 2.1301)	7.5804 (7.0420, 8.1607)	0.8246 (0.4418, 1.3960)
<i>Crest</i>	10.4010 (9.9194, 10.8885)	1.9225 (1.4184, 2.4022)	6.8511 (6.1853, 7.3905)	1.1018 (0.5852, 1.7253)
<i>Mentadent</i>	11.6852 (11.1536, 12.5395)	2.2008 (1.4837, 2.7868)	9.4530 (8.8906, 10.1641)	1.0205 (0.7098, 1.3066)
<i>Liking</i>			0.8619 (0.7376, 0.9938)	0.4973 (0.4012, 0.6006)
<u>Marketing Mix Response</u>				
<i>Price</i>	-4.5587 (-4.8732, -4.2265)	1.9130 (1.5632, 2.2170)	-4.1018 (-4.3216, -3.8930)	1.2670 (1.0696, 1.4795)
<i>Display</i>	0.6307 (0.3005, 0.9779)	0.7452 (0.4706, 1.0122)	0.8189 (0.4941, 1.1506)	0.6568 (0.3742, 0.9861)
<u>Inertia</u>				
<i>Lagged Choice Dummy</i>	2.9772 (2.5472, 3.4862)	1.3591 (0.8821, 1.8160)	2.1879 (1.8942, 2.5097)	0.8434 (0.5620, 1.1301)
<u>Initial Perception Bias</u>				
<i>Arm & Hammer</i>	-3.4873 (-4.5071, -2.4196)		-2.0772 (-3.3359, -0.6711)	
<i>Aim</i>	-2.5850 (-3.8226, -1.4802)		-2.3033 (-3.9876, -0.7866)	
<i>Aquafresh</i>	-2.3066 (-3.0679, -1.5163)		-1.4614 (-2.2674, -0.6588)	
<i>Colgate</i>	-2.0406 (-2.7181, -1.4116)		-0.9605 (-1.6452, -0.3065)	
<i>Crest</i>	-2.0657 (-2.9319, -1.1527)		-0.1101 (-0.8046, 0.5508)	
<i>Mentadent</i>	-3.5958 (-4.8512, -2.5459)		-2.6669 (-3.8979, -1.3811)	
<i>Liking - \overline{Liking}</i>			0.2014 (-0.1264, 0.5694)	
<u>$\text{Log}\left(\frac{1}{\text{Initial Perception Variance}}\right)$</u>				
<i>Intercept</i>	-0.1149 (-0.6124, 0.3626)		-0.0444 (-0.8200, 0.6623)	
<i>Familiarity</i>			0.8153 (0.3414, 1.2878)	
<u>Process Indicator</u>				
<i>Zero-order</i>	0.3120 (0.2710, 0.3529)		0.3949 (0.3533, 0.4353)	
<i>Inertia</i>	0.3137 (0.2768, 0.3503)		0.2846 (0.2514, 0.3192)	
<i>Bayesian Learning</i>	0.3742 (0.3136, 0.3955)		0.3205 (0.2794, 0.3614)	

Notes. Numbers in parenthesis indicate 90% credible set.

[†]Unobserved heterogeneity is measured by the posterior mean of the square root of the diagonal element of V_{Ψ} (Rossi, McCulloch, and Allenby 1996)

Table 4: Concordance Matrix of Assigned Process Indicator

		With Survey			
		Zero-order	Inertia	Bayesian Learning	Sum
Without Survey	Zero-order	0.2542	0.0085	0.0621	0.3249
	Inertia	0.0424	0.2994	0.0282	0.3700
	Bayesian Learning	0.1186	0.0339	0.1525	0.3051
	Sum	0.4153	0.3481	0.2429	

Table 5: Individual Examples of Choice Behavior and Related Estimates

Household #297 (Data)

Purchase Occasion	Choice	Price						
		AH	AM	AF	CG	CR	MT	PS
1	AF	2.4776	1.2100	1.6800	2.5423	2.7414	3.2297	1.4425
2	AF	2.7370	1.4186	1.6800	2.6161	2.5351	4.0740	1.0967
3	AF	2.8332	1.2367	2.4336	2.6964	2.4760	4.2628	1.1080
4	AF	3.2386	1.3800	2.4664	2.5879	2.5579	4.4800	1.0333
5	AF	2.9050	1.4517	2.0600	2.5962	2.4878	3.6792	1.5460
6	AF	4.1200	1.0350	1.6800	2.4572	2.6235	3.2840	1.0320
7	AH	2.4900	1.7150	2.7370	2.5839	2.5066	3.3411	1.7480
	Liking	4	4	7	4	4	1	4
	Familiarity	4	4	7	5	5	2	4
	Mean Price	2.7653	1.425	2.3676	2.5343	2.4377	3.5522	1.3265

Household #27 (Data)

Purchase Occasion	Choice	Price						
		AH	AM	AF	CG	CR	MT	PS
1	CG	1.9162	1.3420	3.1277	2.4098	2.7146	4.4869	1.1900
2	CG	2.9050	1.4517	2.4221	2.5962	2.4878	3.6792	1.5460
3	PS	2.6181	0.9100	2.2073	2.6637	2.7092	3.5250	0.9900
4	AF	2.4520	2.4067	1.8900	2.5972	2.4859	2.9253	1.4617
5	AF	2.9827	1.1050	1.9900	3.3795	2.2432	3.2845	1.5437
6	CR	2.6143	1.7550	2.1663	2.7306	2.4897	3.4651	1.2600
7	CG	2.6143	1.7550	2.1663	2.4900	2.4897	3.4651	1.2600
8	CG	2.9633	1.1400	3.0671	2.5074	3.2077	3.9586	1.4450
9	CR	2.5857	1.2500	2.2689	2.5000	1.9800	3.9062	1.6570
10	AF	3.2989	1.7150	1.8900	2.5839	2.5066	3.3411	1.7480
	Liking	2	1	3	7	6	1	4
	Familiarity	4	2	5	7	5	2	4
	Mean Price	2.7653	1.425	2.3676	2.5343	2.4377	3.5522	1.3265

Household #119 (Data)

Purchase Occasion	Choice	Price						
		AH	AM	AF	CG	CR	MT	PS
1	MT	2.7370	1.4186	2.3378	2.6161	2.5351	2.9700	1.0967
2	CR	2.7027	1.0400	1.8506	2.3820	2.1344	3.9876	1.4200
3	MT	2.6591	1.4200	1.9759	3.1032	2.3555	2.2700	0.9560
4	MT	2.9633	1.1400	3.0671	2.5074	3.2077	2.1800	1.4450
5	CR	3.3194	1.2700	2.3253	2.4308	2.4000	3.5631	1.3525
6	MT	3.2989	1.7150	2.7370	2.5839	2.5066	3.3411	1.7480
	Liking	2	2	2	5	7	7	2
	Familiarity	2	2	2	5	7	7	2
	Mean Price	2.7653	1.425	2.3676	2.5343	2.4377	3.5522	1.3265

Household #297 (Posterior Mean Estimates)

Variable	Level	Without Survey	With Survey
Estimated Proportion of Process Indicator	Zero-order	0.0299	0.7524
	Inertia	0.8263	0.1874
	Bayesian Learning	0.1438	0.0602
Estimated Price Coefficient		-4.6455	-4.101

Household #27 (Posterior Mean Estimates)

Variable	Level	Without Survey	With Survey
Estimated Proportion of Process Indicator	Zero-order	0.1666	0.8017
	Inertia	0.0210	0.0293
	Bayesian Learning	0.8124	0.1690
Estimated Price Coefficient		-6.6956	-5.563

Household #119 (Posterior Mean Estimates)

Variable	Level	Without Survey	With Survey
Estimated Proportion of Process Indicator	Zero-order	0.7377	0.8403
	Inertia	0.0541	0.0377
	Bayesian Learning	0.2082	0.122
Estimated Price Coefficient		-4.0652	-4.4829

Table 6: Counterfactual Sales Promotion Experiments

	Brand						
	Arm & Hammer	Aim	Aquafresh	Colgate	Crest	Mentadent	Pepsodent
I. Free-sampling Experiment							
1) Incremental Own Market Share							
Without Survey	2.89%	0.89%	3.11%	6.15%	6.55%	2.05%	0.09%
With Survey	1.09%	0.34%	1.14%	2.24%	1.37%	0.94%	0.07%
2) Revenue Gains							
Without Survey	174.69\$	26.06\$	184.72\$	404.81\$	420.74\$	173.83\$	2.84\$
With Survey	65.67\$	9.94\$	68.89\$	148.61\$	89.79\$	79.52\$	2.43\$
II. Coupon Experiment							
1) Incremental Own Market Share							
Without Survey	3.01%	1.21%	2.92%	6.63%	6.72%	1.85%	0.71%
With Survey	1.86%	0.61%	1.95%	3.81%	3.76%	1.45%	0.52%
2) Revenue Gains							
Without Survey	131.47\$	18.58\$	115.53\$	326.85\$	338.12\$	119.80\$	10.96\$
With Survey	73.95\$	6.78\$	69.53\$	154.64\$	156.37\$	85.79\$	7.30\$
III. In-store Display Experiment							
1) Incremental Own Market Share							
Without Survey	1.31%	0.08%	0.84%	2.81%	3.08%	0.81%	0.03%
With Survey	1.24%	0.16%	0.89%	1.97%	2.14%	0.71%	0.16%
2) Revenue Gains							
Without Survey	77.76\$	2.64\$	56.05\$	187.76\$	201.22\$	72.86\$	1.03\$
With Survey	71.57\$	4.87\$	54.70\$	129.87\$	139.65\$	60.68\$	5.32\$

Figure 1: Individual Posterior Means of True Mean Quality

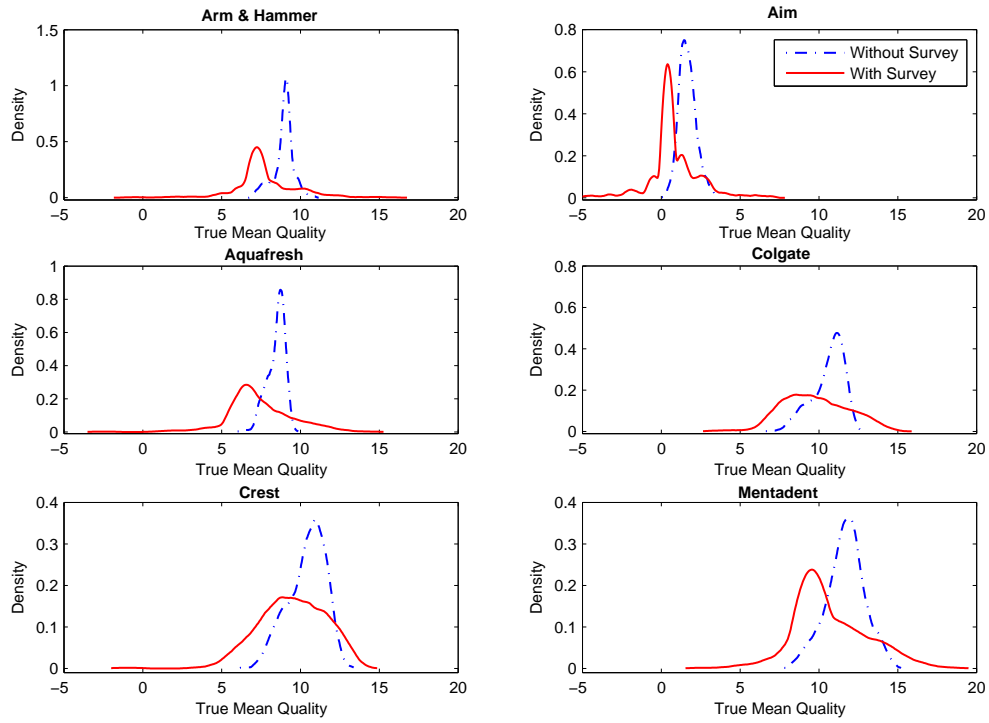


Figure 2: Individual Posterior Means of Marketing Mix Variables

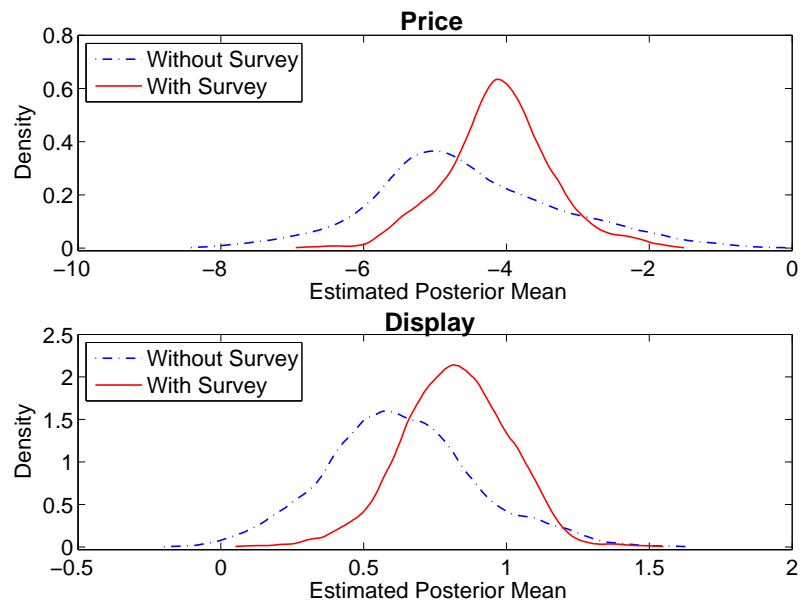


Figure 3: Individual Posterior Means of Inertia Variable

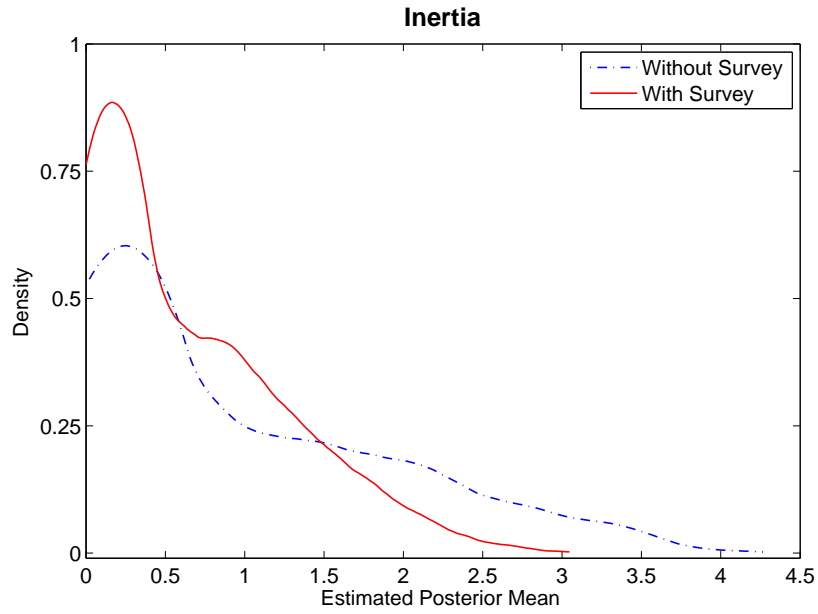


Figure 4: Individual Posterior Means of Initial Perception Bias

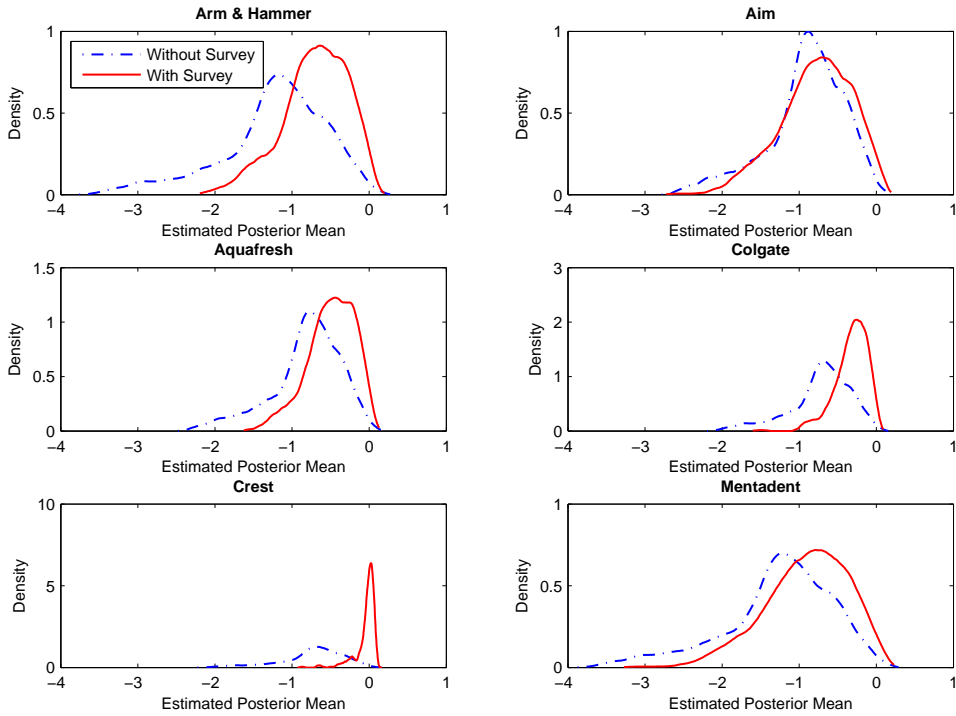


Figure 5: Predicted Aggregate Level Learning

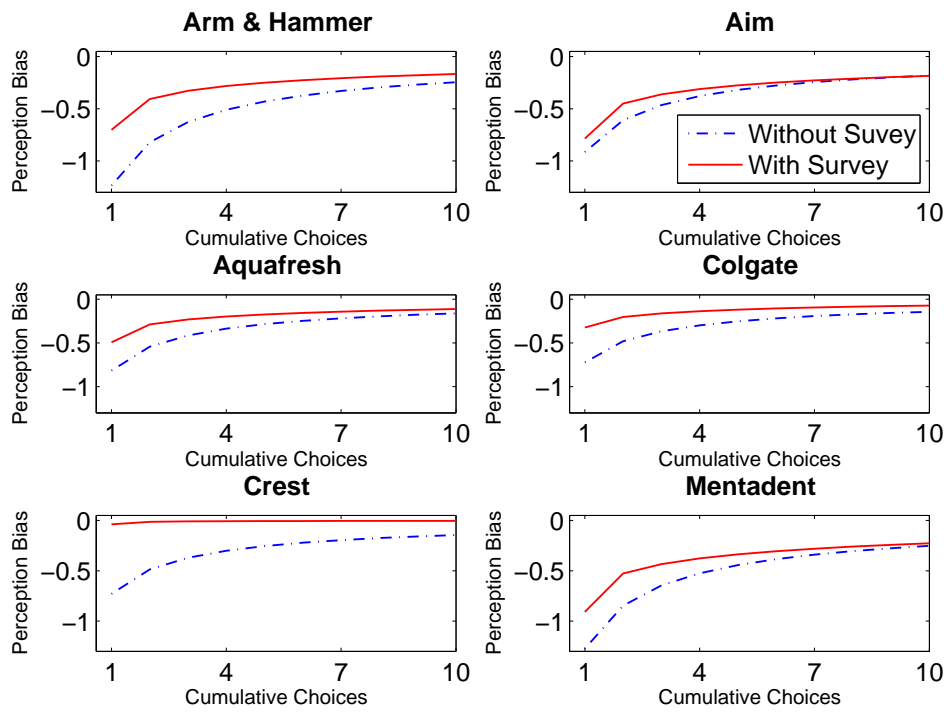


Figure 6: Predicted Individual Level Learning

