RAND Journal of Economics
Vol. 41, No. 3, Autumn 2010
pp. 417–445

State dependence and alternative explanations for consumer inertia

Jean-Pierre Dubé*

Günter J. Hitsch**

and

Peter E. Rossi***

For many consumer packaged goods products, researchers have documented inertia in brand choice, a form of persistence whereby consumers have a higher probability of choosing a product that they have purchased in the past. We show that the finding of inertia is robust to flexible controls for preference heterogeneity and not due to autocorrelated taste shocks. We explore three economic explanations for the observed structural state dependence: preference changes due to past purchases or consumption experiences which induce a form of loyalty, search, and learning. Our data are consistent with loyalty, but not with search or learning. This distinction is important for policy analysis, because the alternative sources of inertia imply qualitative differences in firm's pricing incentives and lead to quantitatively different equilibrium pricing outcomes.

1. Introduction

■ Researchers in both marketing and economics have documented a form of persistence in consumer choice data whereby consumers have a higher probability of choosing products that they have purchased in the past. We call this form of persistence *inertia* in brand choice. Such behavior was first documented by Frank (1962) and Massy (1966); for recent examples, see Keane (1997) and Seetharaman, Ainslie, and Chintagunta (1999). There are two conceptually distinct explanations for the source of inertia in brand choice. One is that past purchases directly influence the consumer's choice probabilities for different brands. Following Heckman (1981), we call this explanation *structural state dependence* in choice. Another explanation is that consumers differ along some serially correlated unobserved propensity to make purchase decisions. Heckman

*University of Chicago and NBER; jdube@chicagobooth.edu.
**University of Chicago; guenter.hitsch@chicagobooth.edu.
***Anderson School of Management, University of California–Los Angeles; perossichi@gmail.com.

We thank Wes Hutchinson, Ariel Pakes, Peter Reiss, and seminar participants at Harvard and UCLA for comments and suggestions. The article has benefited from the comments of two anonymous reviewers and the editor, Phil Haile. We acknowledge the Kilts Center for Marketing at the Booth School of Business, University of Chicago for providing research funds. Dubé was also supported by the Neubauer Family Faculty Fund, and Hitsch was also supported by the Beatrice Foods Company Faculty Research Fund at the Booth School of Business.

(1981) refers to this explanation as *spurious state dependence* because the relationship between past purchases and current choice probabilities only arises if unobserved consumer differences are not properly accounted for. The distinction between the different sources of inertia is important from the point of view of evaluating optimal firm policies such as pricing.

In this article, we document inertia in brand choices using household panel data on purchases of consumer packaged goods (refrigerated orange juice and margarine). We measure inertia using a discrete-choice model that incorporates a consumer's previous brand choice as a covariate. We show that the finding of inertia is robust to controls for unobserved consumer differences, and thus we find evidence for structural state dependence. We then explore three different economic explanations that can give rise to structural state dependence: preference changes due to past purchases (psychological switching costs) which induce a form of loyalty, search, and learning. The patterns in the data are consistent with preference changes, but inconsistent with search and learning.

A standard explanation for the measured inertia is misspecification of the distribution of consumer heterogeneity in preferences. It is difficult to distinguish empirically between structural state dependence and heterogeneity, in particular if the entire set of preference parameters is consumer specific. The extant empirical literature on state dependence in brand choices assumes a normal distribution of heterogeneity.[1] However, a normal distribution may not capture the full extent of heterogeneity. For example, the distribution of brand intercepts could be multimodal, corresponding to different relative brand preferences across groups (or "segments") of consumers. Any misspecification of the distribution of heterogeneity could still lead us to conclude spuriously that consumer choices exhibit structural state dependence. We resolve the potential misspecification problem by using a very flexible, semiparametric heterogeneity specification consisting of a mixture of multivariate normal distributions. We estimate the corresponding choice model using a Bayesian Markov chain Monte Carlo (MCMC) algorithm, which makes inference in a model with such a flexible heterogeneity distribution feasible. We find that past purchases influence current choices, even after controlling for heterogeneity. To confirm that we adequately control for heterogeneity, we reestimate the model based on a reshuffled sequence of brand choices for each household. We no longer find evidence of state dependence with the reshuffled sequence, suggesting there is no remaining unobserved heterogeneity.

A related explanation for inertia is that the choice model errors are autocorrelated, such that a past purchase proxies for a large random utility draw. Following Chamberlain (1985), we first show that past prices predict current choices in a model without a lagged choice variable, which is evidence in favor of structural state dependence. We then exploit the frequent incidence of promotional price discounts in our data. If a past purchase was due to a price discount, the expected random utility draw at the time of the purchase should be smaller than if the purchase was made at a regular price. Therefore, brand choices should exhibit less inertia if the past purchase was initiated by a price discount rather than by a regular price. However, we find no moderating effect of past price discounts on the measured inertia.

Based on our estimates and the tests for unobserved heterogeneity and autocorrelated errors, we conclude that the measured inertia is due to structural state dependence, that is, past choices directly influence current purchase behavior. Unlike most of the past empirical research on inertia in brand choice, we seek to understand the behavioral mechanism that generates structural state dependence. We consider three alternative explanations. A common explanation is that a past purchase or consumption instance alters the current utility derived from the consumption of the product, such that consumers face a form of psychological switching cost in changing brands (Farrell and Klemperer, 2007). We consider this model as our baseline explanation, and refer to this form of structural state dependence as *loyalty* in brand choice. Alternatively, inertia may arise if consumers face search costs and thus do not consider brands which they have not recently

[1] See, for example, Keane (1997), Seetharaman, Ainslie, and Chintagunta (1999), and Osborne (2007). Shum (2004) uses a discrete distribution of heterogeneity.

bought when making a purchase in the product category. To test for a search explanation for inertia, we exploit the availability of in-store display advertising, which reduces search costs. We find that display advertising does not moderate the loyalty effect, and thus conclude that search costs are not the main source of state dependence. Another explanation for structural state dependence is based on consumer learning behavior (see, for example, Osborne, 2007 and Moshkin and Shachar, 2002). A generic implication of learning models is that choice behavior will be nonstationary even when consumers face a stationary store environment. As consumers obtain more experience with the products in a category, the amount of learning declines and their posterior beliefs on product quality converge to a degenerate distribution. Correspondingly, their choice behavior will converge to the predictions from a static brand choice model. On the other hand, if structural state dependence is due to loyalty, there will be no such change in choice behavior over time. We implement a test that exploits this key difference in the predictions of a learning and a loyalty model, and find little evidence in favor of learning. We conclude that the form of structural state dependence in our data is consistent with loyalty, but not with search or learning.

To illustrate the economic significance of distinguishing between inertia as loyalty and spurious state dependence in the form of unobserved heterogeneity or autocorrelated taste shocks, we compare the respective pricing motives and consequences for equilibrium price outcomes. If inertia is due to loyalty, firms can control the evolution of consumer preferences and, thus, face dynamic pricing incentives. In contrast, if inertia is due to unobserved heterogeneity or autocorrelated taste shocks, there are no such dynamic pricing incentives. Thus, the alternative sources of inertia imply qualitative differences in firm's pricing incentives and, as we show using a simulation exercise, also lead to quantitatively different equilibrium pricing outcomes. In a companion article (Dubé, Hitsch, and Rossi, 2009), we provide a detailed analysis of equilibrium pricing if the inertia in brand choice is due to loyalty.

2. Model and econometric specification

■ Our baseline model consists of households making discrete choices among J products in a category and an outside option each time they go to the supermarket. The timing and incidence of trips to the supermarket, indexed by t, are assumed to be exogenous. To capture inertia, we let the previous product choice affect current utilities.

Household h's utility index from product j during the shopping occasion t is

$$u_{jt}^h = \alpha_j^h + \eta^h p_{jt} + \gamma^h \mathbb{I}\{s_t^h = j\} + \epsilon_{jt}^h, \tag{1}$$

where p_{jt} is the product price and ϵ_{jt}^h is a standard iid error term used in most choice models. In the model given by (1), the brand intercepts represent a persistent form of vertical product differentiation that captures the household's intrinsic brand preferences. The household's *state variable* $s_t^h \in \{1, \ldots, J\}$ summarizes the history of past purchases. If a household buys product k during the previous shopping occasion, $t-1$, then $s_t^h = k$. If the household chooses the outside option, then s_t^h remains unchanged: $s_t^h = s_{t-1}^h$. The specification in (1) induces a first-order Markov process on choices. Although the use of the last purchase as a summary of the whole purchase history is frequently used in empirical work, it is not the only possible specification. For example, Seetharaman (2004) considers various distributed lags of past purchases, giving rise to a higher-order Markov process.

If $\gamma^h > 0$, then the model in (1) predicts inertia in brand choices. If a household switches to brand k, the probability of a repeat purchase of brand k is higher than prior to this purchase: the conditional choice probability of repeat purchasing exceeds the marginal choice probability. We refer to γ^h as the *state dependence coefficient*, and we call $\mathbb{I}\{s_t^h = j\}$ the *state dependence term*. To avoid any confusion in our terminology, note that statistical evidence that the state dependence coefficient is positive, $\gamma^h > 0$, need not imply that the brand choices exhibit structural state dependence. Rather, $\gamma^h > 0$ may simply indicate spurious state dependence, for example, if our econometric specification does not fully account for the distribution of preference parameters across consumers or if ϵ^h_{jt} is serially correlated.

□ **Econometric specification.** Assuming that the random utility term, ϵ^h_{jt}, is type I extreme value distributed, household choices are given by a multinomial logit model:

$$\Pr\{j \mid p, s\} = \frac{\exp\left(\alpha^h_j + \eta^h p_j + \gamma^h \mathbb{I}\{s = j\}\right)}{1 + \sum_{k=1}^{J} \exp\left(\alpha^h_k + \eta^h p_k + \gamma^h \mathbb{I}\{s = k\}\right)}. \tag{2}$$

Here, we assume that the mean utility of the outside good is zero, $u_{0t} = 0$.

We denote the vector of household-level parameters by $\theta^h = (\alpha^h_1, \ldots, \alpha^h_J, \eta^h, \gamma^h)$. Preference heterogeneity across household types can be accommodated by assuming that θ^h is drawn from a common distribution. In the extant empirical literature on state-dependent demand, a normal distribution is often assumed, $\theta^h \sim N(\bar{\theta}, V_\theta)$. Frequently, further restrictions are placed on V_θ such as a diagonal structure (see, for example, Osborne, 2007). Other authors restrict the heterogeneity to only a subset of the θ vector. The use of restricted normal models is due, in part, to the limitations of existing methods for estimating random-coefficient logit models.

To allow for a flexible, potentially nonnormal distribution of preference heterogeneity, we employ a Bayesian approach and specify a hierarchical prior with a mixture of normals as the first-stage prior (see, for example, Rossi, Allenby, and McCulloch, 2005). The hierarchical prior provides a convenient way of specifying an informative prior which, in turn, avoids the problem of overfitting even with a large number of normal components. The first stage consists of a mixture of K multivariate normal distributions and the second stage consists of priors on the parameters of the mixture of normals:

$$p(\theta^h \mid \pi, \{\mu_k, \Sigma_k\}) = \sum_{k=1}^{K} \pi_k \phi(\theta^h \mid \mu_k, \Sigma_k) \tag{3}$$

$$\pi, \{\mu_k, \Sigma_k\} \mid b. \tag{4}$$

Here the notation $\cdot \mid \cdot$ indicates a conditional distribution and b represents the hyperparameters of the priors on the mixing probabilities and the parameters governing each mixture component. Mixture-of-normals models are very flexible and can accommodate deviations from normality such as thick tails, skewness, and multimodality.[2]

A useful alternative representation of the model described by (3) and (4) can be obtained by introducing the latent variables $ind_h \in \{1, \ldots, K\}$ that indicate the mixture component from which each consumer's preference parameter vector is drawn:

$$\begin{gathered} \theta^h \mid ind_h, \{\mu_k, \Sigma_k\} \sim \phi(\theta^h \mid \mu_{ind_h}, \Sigma_{ind_h}) \\ ind_h \sim MN(\pi) \\ \pi, \{\mu_k, \Sigma_k\} \mid b. \end{gathered} \tag{5}$$

ind_h is a discrete random variable with outcome probabilities $\pi = (\pi_1, \ldots, \pi_K)$. This representation is precisely that which would be used to simulate data from a mixture of normals, but it is also the same idea used in the MCMC method for Bayesian inference in this model, as detailed in Appendix A. Viewed as a prior, (5) puts positive prior probability on mixtures with different numbers of components, including mixtures with a smaller number of components than K. For example, consider a model that is specified with five components, $K = 5$. *A priori*, there is a

[2] In a separate appendix available upon request, we illustrate this point by simulating data from a model without choice inertia and with a nonnormal distribution of heterogeneity. We find that the normal model for heterogeneity fits a density of the state dependence parameter that is centered away from zero. In contrast, a mixture-of-normals model fits a density that is centered at zero.

positive probability that ind_h takes any of the values $1, \ldots, 5$. *A posteriori*, it is possible that some mixture components are "shut down" in the sense that they have very low probability and are never visited during the navigation of the posterior.

Appendix A provides details on the MCMC algorithm and prior settings used to estimate the mixture-of-normals model (3). We refer the reader to Rossi, Allenby, and McCulloch (2005) for a more thorough discussion.

The MCMC algorithm provides draws of the mixture probabilities as well as the normal component parameters. Thus, each MCMC draw of the mixture parameters provides a draw of the entire multivariate density of household parameters. We can average these densities to provide a Bayes estimate of the household parameter density. We can also construct Bayesian posterior credibility regions[3] for any given density ordinate to gauge the level of uncertainty in the estimation of the household distribution using the simulation draws. That is, for any given ordinate, we can estimate the density of the distribution of either all or a subset of the parameters. A single draw of the ordinate of the marginal density for the ith element of θ can be constructed as follows:

$$p^r_{\theta_i}(\xi) = \sum_{k=1}^{K} \pi^r_k \phi_i\left(\xi \mid \mu^r_k, \Sigma^r_k\right). \tag{6}$$

$\phi_i(\xi \mid \mu_k, \Sigma_k)$ is the univariate marginal density for the ith component of the multivariate normal distribution, $N(\mu_k, \Sigma_k)$.

To obtain a truly nonparametric estimate using the mixture-of-normals model requires that the number of mixture components (K) increases with the sample size (Escobar and West, 1995). Our approach is to fit models with successively larger numbers of components and to gauge the adequacy of the number of components by examining the fitted density as well as the Bayes factor (see the model selection discussion in Section 2) associated with each number of components. What is important to note is that our improved MCMC algorithm is capable of fitting models with a large number of components at relatively low computational cost.

□ **Posterior model probabilities.** To establish that the inertia we observe in the data can be interpreted as structural state dependence, we will compare a variety of different model specifications. Most of the specifications considered will be heterogeneous in that a prior distribution or random-coefficient specification will be assumed for all utility parameters. This poses a problem in model comparison as we are comparing different and heterogeneous models. As a simple example, consider a model with and without the lagged choice term. This is not simply a hypothesis about a given fixed-dimensional parameter, $H_0 : \gamma = 0$, but a hypothesis about a set of household-level parameters. The Bayesian solution to this problem is to compute posterior model probabilities and to compare models on this basis. A posterior model probability is computed by integrating out the set of model parameters to form what is termed the marginal likelihood of the data. Consider the computation of the posterior probability of model M_i:

$$p(M_i \mid D) = \int p(D \mid \Theta, M_i) p(\Theta \mid M_i)\, d\Theta \times p(M_i), \tag{7}$$

where D denotes the observed data, Θ represents the set of model parameters, $p(D \mid \Theta, M_i)$ is the likelihood of the data for M_i, and $p(M_i)$ is the prior probability of model i. The first term in (7) is the marginal likelihood for M_i.

$$p(D \mid M_i) = \int p(D \mid \Theta, M_i) p(\Theta \mid M_i)\, d\Theta \tag{8}$$

[3] The Bayesian posterior credibility region is the Bayesian analogue of a confidence interval. The 95% posterior credibility region is an interval which has .95 probability under the posterior. We compute equal-tailed estimates of the posterior credibility region by using quantiles from the MCMC draws.

The marginal likelihood can be computed by reusing the simulation draws for all model parameters that are generated by the MCMC algorithm using the method of Newton and Raftery (1994).

$$\hat{p}(D \mid M_i) = \left(\frac{1}{R} \sum_{r=1}^{R} \frac{1}{p(D \mid \Theta^r, M_i)} \right)^{-1} \tag{9}$$

$p(D \mid \Theta, M_i)$ is the likelihood of the entire panel for model i. In order to minimize overflow problems, we report the log of the trimmed Newton-Raftery MCMC estimate of the marginal likelihood. Assuming equal prior model probabilities, Bayesian model comparison can be done on the basis of the marginal likelihood (assuming equal prior model probabilities).

Posterior model probabilities can be shown to have an automatic adjustment for the effective parameter dimension. That is, larger models do not automatically have higher marginal likelihood, as the dimension of the problem is one aspect of the prior that always matters. Although we do not use asymptotic approximations to the posterior model probabilities, the asymptotic approximation to the marginal likelihood illustrates the implicit penalty for larger models (see, for example, Rossi, McCulloch, and Allenby, 1996).

$$log(p(D \mid M_i)) \approx log(p(D \mid \hat{\Theta}_{MLE}, M_i)) - \frac{p_i}{2} log(n) \tag{10}$$

p_i is the effective parameter size for M_i and n is the sample size. Thus, a model with the same fit or likelihood value but a larger number of parameters will be "penalized" in marginal likelihood terms. Choosing models on the basis of the marginal likelihood can be shown to be consistent in model selection in the sense that the true model will be selected with probability converging to one as the sample size becomes infinite (e.g., Dawid, 1992).

3. Data

■ We estimate the logit demand model described above using household panel data containing information on purchases in the refrigerated orange juice and the 16 ounce tub margarine consumer packaged goods categories. The panel data were collected by AC Nielsen for 2100 households in a large Midwestern city between 1993 and 1995. In each category, we focus only on those households that purchase a brand at least twice during our sample period. We use 355 households to estimate orange juice demand and 429 households to estimate margarine demand. We also use AC Nielsen's store-level data for the same market to obtain the weekly prices and point-of-purchase marketing variables for each of the products that were not purchased on a given shopping trip.

Table 1 lists the products considered in each category as well as the product purchase and no-purchase shares and average prices. We define the outside good in each category as follows. In the refrigerated orange juice category, we define the outside good as any fresh or canned juice product purchase other than the brands of orange juice considered. In the tub margarine category, we define the outside good as any trip during which another margarine or butter product was purchased.[4] In Table 1, we see a no-purchase share of 23.8% in refrigerated orange juice and 40.8% in tub margarine. Using these definitions of the outside good, we model only those shopping trips where purchases in the product category are considered.

In our econometric specification, we will be careful to control for heterogeneity as flexibly as possible to avoid confounding structural state dependence with unobserved heterogeneity. Even with these controls in place, it is still important to ask which patterns in our consumer shopping panel help us to identify state dependence effects. In Table 2, we show that the marginal purchase probability is considerably smaller than the conditional repurchase probability for each of the products considered. Thus, we observe inertia in the raw data. However, the raw data alone are inadequate to distinguish between structural state dependence and unobserved heterogeneity in

[4] Although not reported, our findings in the margarine category are qualitatively similar if we use a broader definition of the outside option based on any spreadable product (jams, jellies, margarine, butter, peanut butter, etc.).

TABLE 1 Data Description

Product	Average Price ($)	Trips (%)
	Margarine	
Promise	1.69	14.3
Parkay	1.63	5.4
Shedd's	1.07	13.8
I Can't Believe It's Not Butter!	1.55	25.6
No purchase		40.8
No. of households	429	
No. of trips per household	16.7	
No. of purchases per household	9.9	

Product	Average Price ($)	Trips (%)
	Refrigerated orange juice	
64 oz Minute Maid	2.21	11.1
Premium 64 oz Minute Maid	2.62	7.0
96 oz Minute Maid	3.41	14.7
64 oz Tropicana	2.26	6.7
Premium 64 oz Tropicana	2.73	28.8
Premium 96 oz Tropicana	4.27	8.0
No purchase		23.8
No. of households	355	
No. of trips per household	12.3	
No. of purchases per household	9.4	

TABLE 2 Repurchase Rates

Product	Purchase Frequency	Repurchase Frequency	Repurchase Frequency after Discount
		Margarine	
Promise	.24	.83	.85
Parkay	.09	.90	.86
Shedd's	.23	.81	.80
ICBINB	.43	.88	.88
		Refrigerated orange juice	
Minute Maid	.43	.78	.74
Tropicana	.57	.86	.83

consumer tastes. The identification of state dependence in our context is aided by the frequent temporary price changes typically observed in supermarket scanner data. If there is sufficient price variation, we will observe consumers switching away from their preferred products. The detection of state dependence relies on spells during which the consumer purchases these less-preferred alternatives on successive visits, even after prices return to their "typical" levels.

We use the orange juice category to illustrate the source of identification of state dependence in our data. We classify each product's weekly price as either "regular" or "discount," where the latter implies a temporary price decrease of at least 5%. Conditional on a purchase, we observe 1889 repeat purchases (spells) out of our total 3328 purchases in the category. In many cases, the spell is initiated by a discount price. For instance, nearly 60% of the cases where a household purchases something other than its favorite brand, the product chosen is offering a temporary price discount. We compare the repeat-purchase rate for spells initiated by a price discount (i.e., a household repeat buys a product that was on discount when previously purchased) to the marginal probability of a purchase in Table 2. In this manner, the initial switch may not merely reflect heterogeneity in tastes. For all brands of Minute Maid orange juice, the sample repurchase probability conditional on a purchase initiated by a discount is .74, which exceeds the marginal purchase probability of .43. The same is true for Tropicana brand products, with the conditional repurchase probability of .83 compared to the marginal purchase probability of .57. These patterns are suggestive of a structural relationship between current and past purchase behavior, as opposed to persistent, household-specific differences.

Inertia in brand choices is one possible form of dependence in shopping behavior over time. Another frequently cited source of non-zero-order purchase behavior is household inventory holdings or stockpiling (see, for example, Erdem, Imai, and Keane, 2003 and Hendel and Nevo, 2006). Households that engage in stockpiling change the timing of their purchases and the quantities they purchase (i.e., pantry loading) based on their current product inventories, current prices, and their expectations of future price changes. Although stockpiling has clear implications for purchase timing, it does not predict a link between past and current brand choices, that is, inertia.

4. Inertia as structural state dependence versus spurious state dependence

□ **Heterogeneity and state dependence.** It is well known that structural state dependence and heterogeneity can be confounded (Heckman, 1981). We have argued that frequent price discounts or sales provide a source of brand switching that allows us to separate structural state dependence in choices from persistent heterogeneity in household preferences. However, it is an empirical question as to whether or not state dependence is an important force in our data. With a normal distribution of heterogeneity, a number of authors have documented that positive state dependence is present in consumer packaged goods panel data (see, for example, Seetharaman, Ainslie, and Chintagunta, 1999). However, there is still the possibility that these results are not robust to controls for heterogeneity using a flexible or nonparametric distribution of preferences. Our approach consists of fitting models with and without an inertia term and with various forms of heterogeneity. Our mixture-of-normals approach nests the normal model in the literature.

Table 3 provides log marginal likelihood results that facilitate assessment of the statistical importance of heterogeneity and state dependence. All log marginal likelihoods are estimated using a Newton-Raftery-style estimator that has been trimmed of the top and bottom 1% of likelihood values as is recommended in the literature (Lopes and Gamerman 2006). We compare models without heterogeneity to a normal model (a one-component mixture) and to a five-component mixture model.

As is often the case with consumer panel data (Allenby and Rossi, 1999), there is pronounced heterogeneity. Recall that if two models have equal prior probability, the difference in log marginal likelihoods is related to the ratio of posterior model probabilities:

$$log\left(\frac{p(M_1 \mid D)}{p(M_2 \mid D)}\right) = log(p(D \mid M_1)) - log(p(D \mid M_2)). \tag{11}$$

Table 3 shows that in a model specification including a state dependence term the introduction of normal heterogeneity improves the log marginal likelihood by about 2700 points for margarine and about 1700 points for refrigerated orange juice. Therefore, the ratio of posterior probabilities is about exp(2700) in margarine and about exp(1700) in orange juice, providing overwhelming evidence in favor of a model with heterogeneity in both product categories.

The normal model of heterogeneity does not appear to be adequate for our data, as the log marginal likelihood improves substantially when a five-component mixture model is used. In a model including a state dependence term, moving from one to five normal components increases the log marginal likelihood by 45 points (from −4906 to −4861) for margarine products and 52 points for orange juice. Remember that the Bayesian approach automatically adjusts for effective

TABLE 3 Log Marginal Likelihood Values for Different Model Specifications

	Margarine	Orange Juice
Models not allowing for state dependence		
Homogeneous model	−10211	−7612
Five normal components	−4922	−4528
Five normal components, lagged prices	−4829	−4389
Models allowing for state dependence		
Homogeneous model	−7618	−6297
One normal component	−4906	−4486
Five normal components	−4861	−4434
Randomized purchase sequence, five normal components (30 replications)		
Median	−4908	−4501
2.5th percentile	−4924	−4533
97.5th percentile	−4885	−4470
Interaction with discount, five normal components	−4854	−4419
Brand-specific inertia, five normal components	−4822	−4364
Learning models, five normal components		
Experienced shopper dummy	−4884	−4477
Main effect of brand experience	−4654	−4297
Main and interaction effects of brand experience	−4620	−4293

Note: The models allowing for state dependence include the state variable indicating the last purchased product as a covariate, whereas the models not allowing for state dependence impose the restriction $\gamma^h = 0$.

parameter size (see Section 2) such that the differences in log marginal likelihoods documented in Table 3 represent a meaningful improvement in model fit.

Figures 1–4 illustrate the importance of a flexible heterogeneity distribution. Each figure plots the estimated marginal distribution of the intercept, price,[5] and state dependence coefficients[6] from the five-component mixture model (we display the posterior mean as the Bayes estimate of each density value). The shaded envelope enclosing the marginal densities is a 90% pointwise posterior credibility region. The graphs also display the corresponding coefficient distributions from a one-component model of heterogeneity. Several of the parameters exhibit a dramatic departure from normality. For example, in the margarine category, the Shedd's and Parkay brand intercepts (Figure 1) have a noticeably bimodal marginal distribution. For the Shedd's brand, one mode is centered on a positive value, indicating a strong brand preference for Shedd's. The other mode is centered on a negative value, reflecting consumers who view Shedd's as inferior to the outside good. There is a similar bimodality in the orange juice intercepts displayed in Figure 3. The price coefficients (Figures 2 and 4) are also nonnormal, exhibiting pronounced bimodality in the margarine category and left skewness in the orange juice category.

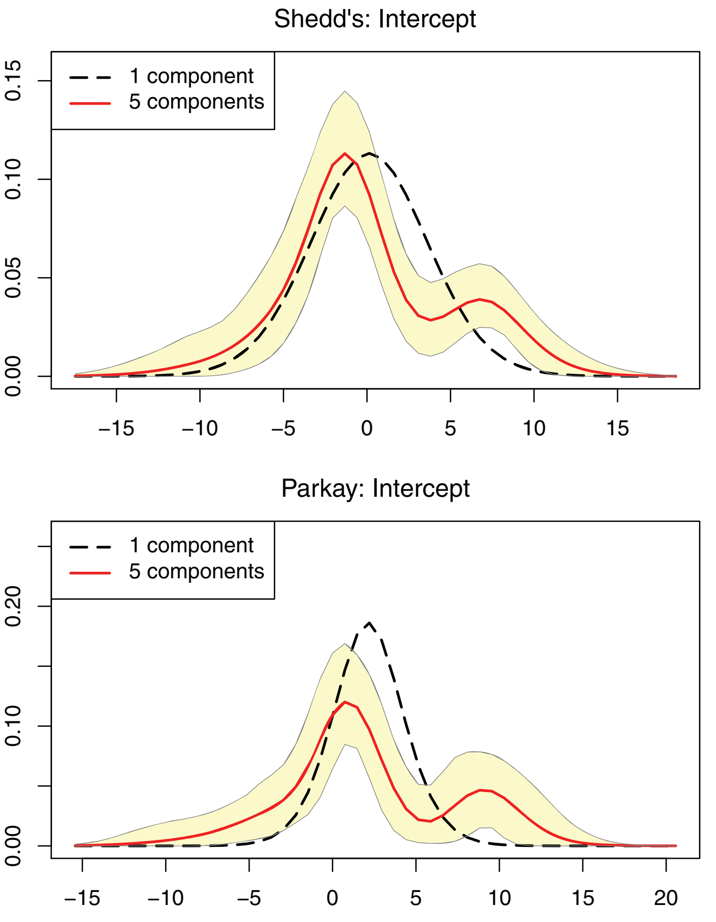
Thus, in our data, the findings indicate that there is good reason to doubt the appropriateness of the standard normal assumption for many of the choice model parameters. This opens the possibility that the findings in the previous literature documenting structural state dependence are influenced, at least in part, by incorrect distributional assumptions. However, in our data, we find evidence for state dependence even when a flexible five-component normal model of heterogeneity is specified. The log marginal likelihood increases from −4922 to −4861 when

[5] A potential concern is that we do not constrain the price coefficient to be negative and, accordingly, the population-level marginal distribution for this coefficient places mass on positive values. However, when we compute the posterior mean price coefficients for each household, we get a positive coefficient in only 10 of the 429 cases (2%) for margarine, and 5 of the 355 cases (1%) for orange juice.

[6] The fitted density of the state dependence coefficient, although centered above zero, does place some mass on negative values in both categories. If we compute each household's posterior mean coefficient, we find a negative value in 98 of the 429 cases for margarine, and in 28 of the 355 cases for orange juice. One interpretation of the negative coefficient is that some households seek variety in their brand choices over time.

FIGURE 1

DISTRIBUTION OF BRAND INTERCEPTS: MARGARINE



The graphs display the pointwise posterior mean and 90% credibility region of the marginal density of margarine brand intercepts (α_j^h). The results are based on a five-component mixture-of-normals heterogeneity specification. For comparison purposes, we also show the results from a one-component heterogeneity specification.

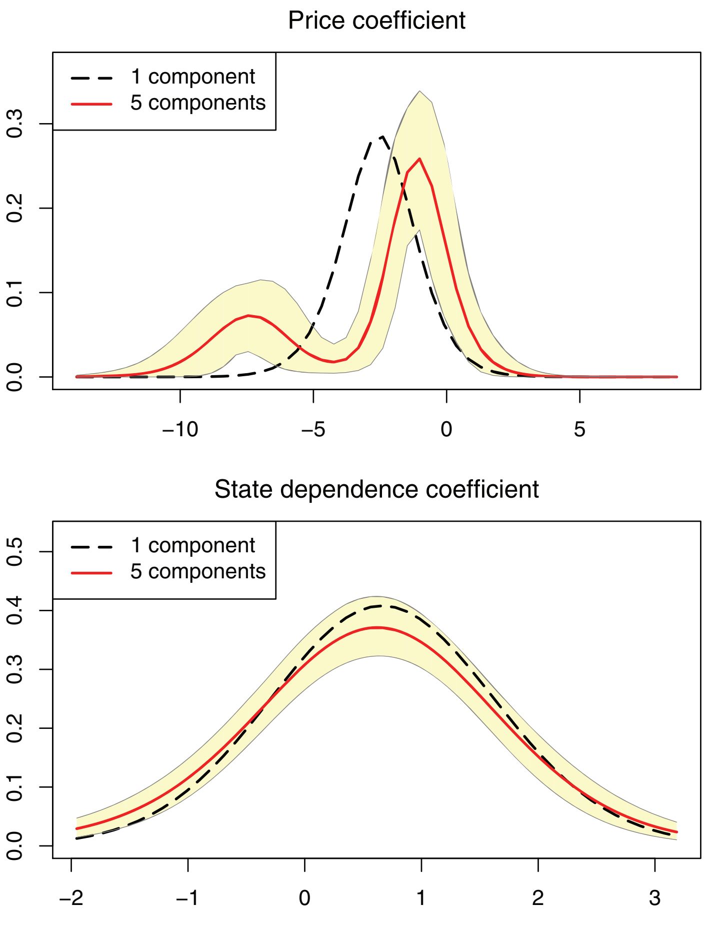
a state dependence term is added to a five-component model for margarine and from −4528 to −4434 for refrigerated orange juice. Although not definitive evidence, this result suggests that the findings of state dependence in the literature are not artifacts of the normality assumption commonly used. Figures 2 and 4 show the marginal distribution of the state dependence parameter, which is well approximated by a normal distribution for these two product categories.

The five-component normal mixture is a very flexible model for the joint density of the choice model parameters. However, before we can make a more generic "semiparametric" claim that our results are not dependent on the functional form of the heterogeneity distribution, we must provide evidence of the adequacy of the five-component model. Our approach is to fit a ten-component model, which is a very highly parameterized specification. In the margarine category, for example, the ten-component model has 449 parameters (the coefficient vector is eight-dimensional[7]). Although not reported in the tables, the log marginal likelihood remains nearly unchanged as we move from five to ten components: from −4843 to −4842 for margarine

[7] There are $36 \times 10 = 360$ unique variance-covariance parameters plus $10 \times 8 = 80$ mean parameters plus 9 mixture probabilities.

FIGURE 2

DISTRIBUTION OF PRICE AND STATE DEPENDENCE COEFFICIENTS: MARGARINE



The graphs display the pointwise posterior mean and 90% credibility region of the marginal density of the margarine price coefficient (η^h) and state dependence coefficient (γ^h). The results are based on a five-component mixture-of-normals heterogeneity specification. For comparison purposes, we also show the results from a one-component heterogeneity specification.

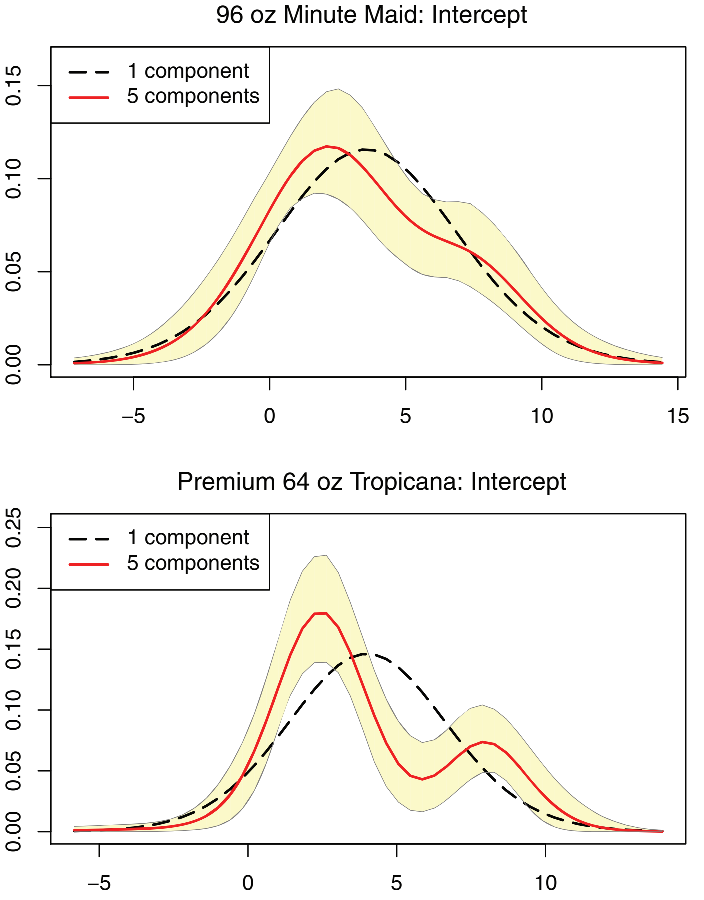
and −4434 to −4435 for orange juice. These results indicate no value from increasing the model flexibility beyond five components. The posterior model probability results and the high flexibility of the models under consideration justify the conclusion that we have accommodated heterogeneity of an unknown form.

□ Robustness checks

State dependence or a misspecified distribution of heterogeneity? We perform a simple additional check to test for the possibility that the lagged choice coefficient proxies for a misspecification of the distribution of heterogeneity. Suppose there is no structural state dependence and that the coefficient on the lagged choice picks up taste differences across households that are not accounted for by the assumed functional form of heterogeneity. Then, if we randomly reshuffle the order of shopping trips, the coefficient on the lagged choice will not change and still provide misleading evidence for state dependence. In Table 3, we show the median, 2.5th percentile, and 97.5th percentile values of the log marginal likelihoods for a five-component model with a

FIGURE 3

DISTRIBUTION OF BRAND INTERCEPTS: REFRIGERATED ORANGE JUICE



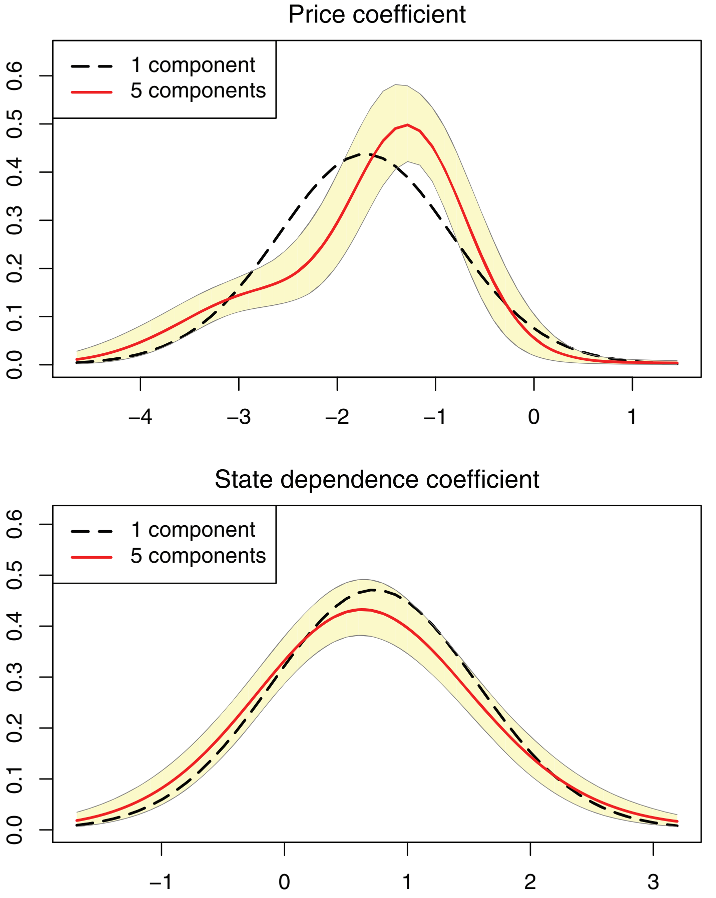
The graphs display the pointwise posterior mean and 90% credibility region of the marginal density of refrigerated orange juice brand intercepts (α_j^h). The results are based on a five-component mixture-of-normals heterogeneity specification. For comparison purposes, we also show the results from a one-component heterogeneity specification.

state dependence term, which we fitted to 30 randomly reshuffled purchase sequences. A 95% interval of the log marginal likelihoods based on the reshuffled purchase sequences contains the log marginal likelihood pertaining to the model that does not include a state dependence term. Furthermore, the 95% interval is strictly below the marginal likelihood pertaining to the model that includes a state dependence term based on the correct choice sequence. We thus find additional strong evidence against the possibility that the lagged choice proxies for a misspecified heterogeneity distribution.

State dependence or autocorrelation? Although the randomized sequence test gives us confidence that we have found evidence of a non-zero-order choice process, it does not help us to distinguish between structural state dependence and a model with autocorrelated choice errors. If the choice model errors are autocorrelated, a past purchase will proxy for a large past and hence also a large current random utility draw. Thus, a past purchase will predict current choice behavior. A model with both state dependence and autocorrelated errors is considered in Keane (1997). Using a normal distribution of heterogeneity and a different estimation method, he finds that the estimated degree of state dependence remains largely unchanged if autocorrelated random utility

FIGURE 4

DISTRIBUTION OF PRICE AND STATE DEPENDENCE COEFFICIENTS: REFRIGERATED ORANGE JUICE



The graphs display the pointwise posterior mean and 90% credibility region of the marginal density of the refrigerated orange juice price coefficient (η^h) and state dependence coefficient (γ^h). The results are based on a five-component mixture-of-normals heterogeneity specification. For comparison purposes, we also show the results from a one-component heterogeneity specification.

terms are allowed for. The economic implications of the two models are markedly different. If state dependence represents a form of state-dependent utility or loyalty, firms can influence the loyalty state of the customer, and this has, for example, long-run pricing implications. However, the autocorrelated errors model does not allow for interventions to induce loyalty to a specific brand. We will discuss these points in Section 6.

In order to distinguish between a model with a state dependence term and a model with autocorrelated errors, we implement the suggestion of Chamberlain (1985). We consider a model with a five-component normal mixture for heterogeneity, no state dependence term, but including lagged prices defined as the prices at the last purchase occasion. In a model with structural state dependence, the product price can influence the consumer's state variable and this will affect subsequent choices. In contrast, in a model with autocorrelated errors, prices do not influence the persistence in choices over time. In Table 3, we compare the log marginal likelihood of a model without a state dependence term and a five-component normal mixture with the log marginal likelihood of the same model including lagged prices. The addition of lagged prices improves the log marginal likelihood by 93 points for margarine and by 139 points for refrigerated orange juice. This is strong evidence in favor of structural state dependence specification over autocorrelation in random utility terms.

A limitation of the Chamberlain suggestion (as noted by both Chamberlain himself and Erdem and Sun, 2001) is that consumer expectations regarding prices (and other right-hand-side variables) might influence current choice decisions. Lagged prices might simply proxy for expectations even in the absence of structural state dependence. Thus, the importance of lagged prices as measured by the log marginal likelihood is suggestive but not definitive.

As another comparison between a model with autocorrelated errors and a model with structural state dependence, we exploit the price discounts or sales in our data. As autocorrelated errors are independent across households and independent of the price discounts, we can differentiate between state-dependent and autocorrelated error models by examining the impact of price discounts on measured state dependence. Suppose that a household chooses product j at shopping occasion t, denoted by $d_{jt} = 1$. ϵ_{jt} is the random utility term of product j, which may be autocorrelated or independent across time. Given that the household chooses product j and given *any* price vector p_t, the random utility term ϵ_{jt} must be larger than the population average, $\mathbb{E}(\epsilon_{jt} \mid p_t, d_{jt} = 1) > \mathbb{E}(\epsilon_{jt})$. Therefore, under autocorrelation it is also true that $\mathbb{E}(\epsilon_{j\tau} \mid p_\tau, d_{jt} = 1) > \mathbb{E}(\epsilon_{j\tau})$ at a subsequent shopping occasion $\tau > t$ and, hence, we would find spurious state dependence if we included lagged choices in the choice model. Our test for autocorrelation exploits the fact that, if the incidence of price discounts is independent across products,[8] $\mathbb{E}(\epsilon_{jt} \mid p_j^R, d_{jt} = 1) > \mathbb{E}(\epsilon_{jt} \mid p_j^D, d_{jt} = 1)$, where p_j^R is a regular price and $p_j^D < p_j^R$ is a discounted price. For type I extreme value distributed random utility terms, this follows from the expression $\mathbb{E}(\epsilon_j \mid p, d_j = 1) = e - \log(\Pr\{j \mid p\})$, where e is Euler's constant, and Appendix B shows that the statement is also true for more general distributions of ϵ. Therefore, under autocorrelated random utility terms, the correlation between the past purchase state and the current product choice should be lower if the loyalty state was initiated by a price discount rather than by a regular price.

To implement this test, we estimate the following model:

$$u_{jt} = \alpha_j + \eta_j p_{jt} + \gamma_1 \mathbb{I}\{s_t = j\} + \gamma_2 \mathbb{I}\{s_t = j\} \cdot \mathbb{I}\{discount_{s_t} = j\} + \epsilon_{jt}. \tag{12}$$

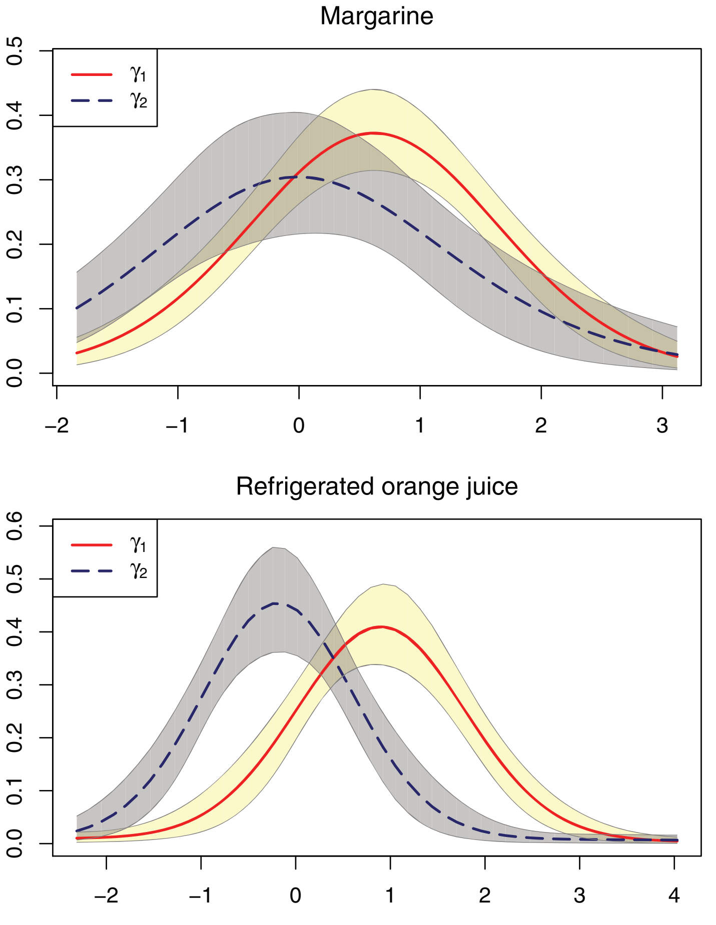
The term $discount_{s_t}$ indicates whether the brand corresponding to the customer's current state was on discount when it was last purchased. In a model with autocorrelated errors, the magnitude of the spurious state dependence effect should be lower for states generated by discounts, that is, $\gamma_2 < 0$. On the other hand, if the errors are independent across time and if the past product choice directly affects the current purchase probability for the same brand, then $\gamma_1 > 0$ and $\gamma_2 = 0$.

Table 3 reports the log marginal likelihood for model (12). Adding the interaction with the discount variable to the original five-component model improves the model fit by a modest 7 points for margarine and 15 points for orange juice. Figure 5 displays the fitted marginal distribution of the state dependence parameter, γ_1, and the interaction term of state dependence with price discounts, γ_2. Recall that we allow for an entire distribution of parameters across the population of consumers so that we cannot provide the Bayesian analogue of a point estimate and a confidence interval for γ_1 and γ_2. The distribution of the main effect of state dependence, γ_1, is centered at a positive value for both categories. Also, comparing Figure 5 to Figures 2 and 4, we see that the estimated distribution of γ_1 changes little if the additional interaction term is included in the model. The distribution of γ_2 is centered at zero for margarine. For orange juice, γ_2 is centered on a slightly negative value; however, the 95% posterior credibility region of the population mean of γ_2 contains zero. Combining the evidence from this test with the results from the Chamberlain test reported above, we conclude that overall there is scant evidence that the measured state dependence is due to autocorrelated errors.

[8] We rarely see more than one brand in a category on sale at the same time. In the margarine category, for instance, less than 2% of the trips have two or more products on sale at the same time.

FIGURE 5

TESTING FOR AUTOCORRELATION



The graphs display the pointwise posterior mean and 90% credibility region of the marginal density of the coefficients γ_1 and γ_2 in model (12). γ_1 is the main state dependence coefficient, and γ_2 represents the effect of the interaction between the purchase state and the presence of a price discount when the product was last purchased. We expect that $\gamma_2 < 0$ under autocorrelated taste shocks. The results are based on a five-component mixture-of-normals heterogeneity specification.

□ **Fixed store effects and price endogeneity.** As in much of the demand estimation literature, the potential endogeneity of supply-side variables could bias our parameter estimates. A bias toward zero in the estimated price coefficient could also spuriously indicate state dependence. For instance, if a consumer begins purchasing a product repeatedly due to low prices and the price parameter is underestimated, this behavior could be misattributed to state dependence. In our current context, we pool trips across 40 stores in the two largest supermarket chains in the market. It is possible that unobserved (to the researcher) store-specific factors, such as shelf space and/or store configuration, could differentially influence a consumer's propensity to purchase across stores. Endogeneity bias might arise if retailers condition on these store-level factors when they set their prices, creating a correlation between the observed shelf prices and the unobserved store effects. Empirically, most of the price variation in our data is across brands, a dimension that we control for with brand intercepts in the choice model. Thus, although endogeneity is a possibility, in our data only 2% of the variance in prices is explained by store effects, and only 1% is explained by chain effects.

To control for this potential source of endogeneity, we reestimate demand with a complete set of store-specific intercepts. Household h's utility index from product j during shopping occasion t at store k is

$$u^h_{jtk} = \alpha^h_j + \eta^h p_{jt} + \gamma^h \mathbb{I}\{s^h_t = j\} + \xi_k + \epsilon^h_{jt}, \tag{13}$$

where ξ_k is common across all consumers and shopping occasions. ξ_k does not enter the utility of the outside good. For estimation, we assume the following prior structure on each ξ_k:

$$\xi_k \sim N\left(\bar{\xi}, A_\xi^{-1}\right).$$

We use the prior settings $\bar{\xi} = 0$ and $A_\xi^{-1} = .01$.

We fit the state dependence model with fixed store effects to the margarine data. This model places a heavy burden on our estimator, as it adds 38 additional parameters.[9] Although, store effects improve fit substantially for the homogeneous specification (the marginal likelihood increases from −7618 to −7494), the improvement in fit is modest for the heterogeneous, five-component specification (the marginal likelihood increases from −4861 to −4853).

In Figure 6 , we plot the price and state dependence coefficients for a five-component mixture-of-normals specification both with and without controls for store effects. The fitted density for state dependence is identical in the two cases. The fitted density for the price coefficient looks different if store effects are included in the model (e.g., unimodal as opposed to bimodal). However, these differences may simply be due to sampling error, a factor we can assess by noting the high degree of overlap in the 95% posterior credibility regions. In summary, our main finding is that our estimates of structural state dependence are not affected by price endogeneity due to unobserved, store-specific effects.

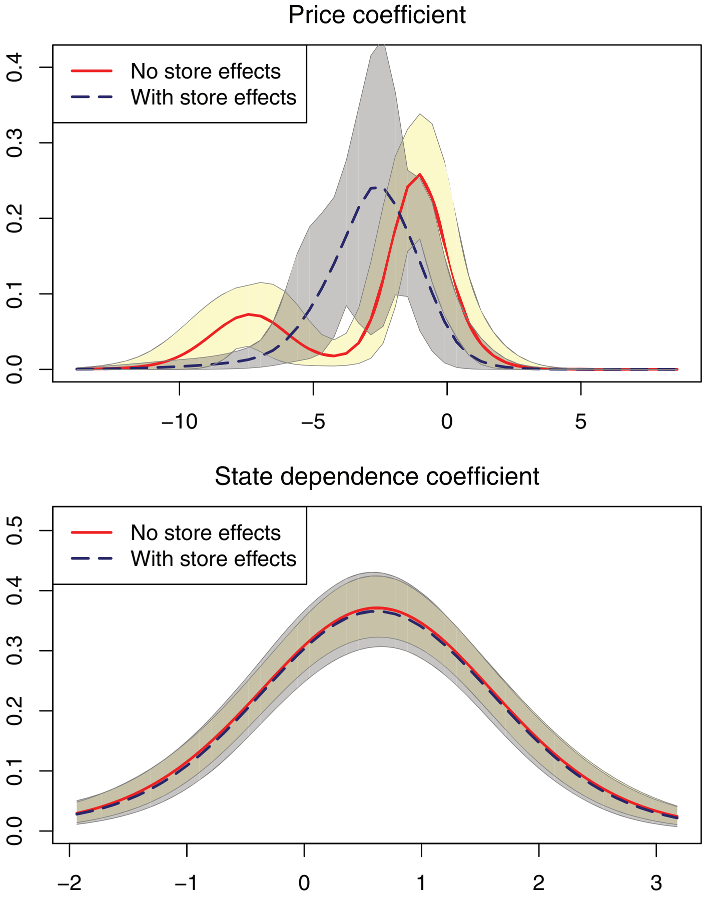
□ **Brand-specific state dependence.** In the basic utility specification (1), state dependence is captured by a parameter that is constrained to be identical across brands. Several authors have found the measurement of state dependence to be difficult (see, for example, Keane, 1997; Seetharaman, Ainslie, and Chintagunta, 1999; Erdem and Sun, 2001) even with a one-component normal model for heterogeneity. The reason for imposing one state dependence parameter could simply be a need for greater efficiency in estimation. However, it would be misleading to report state dependence effects if these are limited to, for example, only one brand in a set of products. It also might be expected that some brands with unique packaging or trademarks might display more state dependence than others. It is also possible that the formulation of some products may induce more state dependence via some mild form of "addiction" in that some tastes are more habit forming than others. For these reasons, we consider an alternative formulation of the model with brand-specific state dependence parameters. Our Bayesian methods have a natural advantage for highly parameterized models in the sense that if a model is weakly identified from the data, the prior keeps the posterior well defined and regular.

A five-component mixture-of-normals model with brand-specific state dependence fits the data with a higher log marginal likelihood for both categories. The log marginal likelihood increases from −4861 to −4822 for margarine and −4434 to −4364 for orange juice. However, there is a difference between substantive and statistical significance. For this reason, we plot the fitted marginal densities for the state dependence parameters for each brand in Figures 7 and 8 and compare them to the state dependence distributions from the baseline model. In the margarine category, all four distributions are centered close to the baseline, constrained specification. In the orange juice category, the two largest 96 ounce brands shown have higher inertia than the two largest 64 ounce brands. The prior distribution on the state dependence parameters is centered

[9] We pool three of the stores in the smaller chain into one group, as none of them has more than 20 observed trips in our data.

FIGURE 6

DISTRIBUTION OF PRICE AND STATE DEPENDENCE COEFFICIENTS WITH AND WITHOUT CONTROLLING FOR STORE EFFECTS: MARGARINE



The graphs display the pointwise posterior mean and 90% credibility region of the marginal density of the margarine price coefficient (η^h) and state dependence coefficient (γ^h). The results are based on a five-component mixture-of-normals heterogeneity specification and are shown for model specifications with and without store effects.

at zero and very diffuse.[10] This means that the data have moved us to a posterior which is much tighter than the prior and moved the center of mass away from zero. Thus, our results are not simply due to the prior specification but are the result of evidence in our data.

The main conclusion is that allowing for brand-specific state dependence parameters does not reduce the importance of state dependence nor restrict these effects to a small subset of brands.

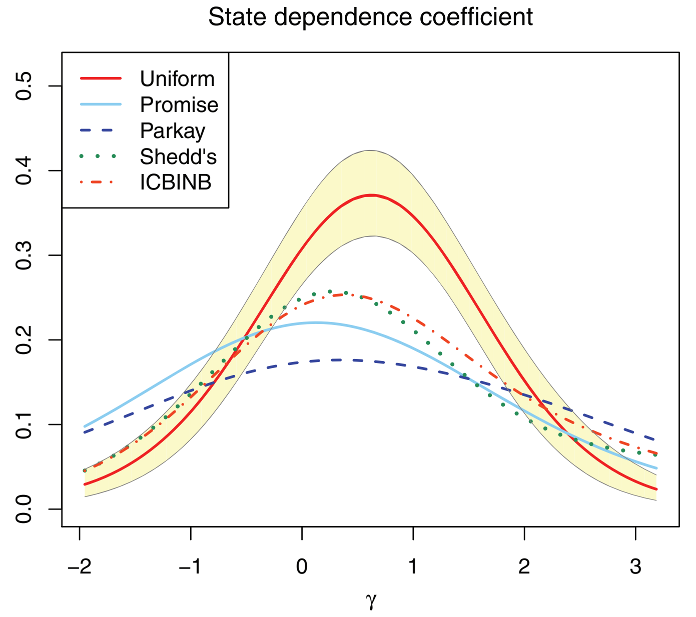
5. Alternative sources of structural state dependence

■ We have found evidence for structural state dependence in brand choice even after controlling for a very flexible distribution of preference heterogeneity. The estimated state dependence effects

[10] It should be noted that, as detailed in Appendix A, our prior is a prior on the parameters of the mixture of normals—the mixing probabilities and each component mean vector and covariance matrix. This induces a prior on the distribution over parameters and the resultant marginal densities. Although this is of no known analytic form, the fact that our priors on each component parameter are diffuse mean that the prior on the distributions is also diffuse.

FIGURE 7

DISTRIBUTION OF BRAND-SPECIFIC STATE DEPENDENCE COEFFICIENTS: MARGARINE



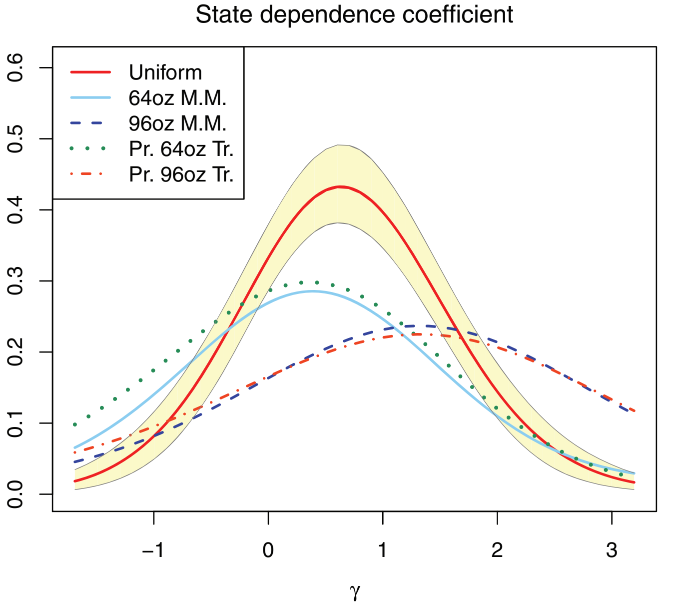
The graph displays the pointwise posterior mean and 90% credibility region of the marginal density of the state dependence coefficient (γ^h), based on a five-component mixture-of-normals heterogeneity specification. We show the densities both for a model specification with a uniform (across-brands) state dependence coefficient and for a specification allowing for brand-specific state dependence coefficients.

are unlikely to be the result of autocorrelated random utility shocks. In this section, we explore different behavioral mechanisms that could give rise to the structural state dependence effects observed in the data. Our baseline explanation is that a past purchase or consumption of a brand directly changes a consumer's preference for the brand. We refer to this form of structural state dependence as *loyalty*. Such loyalty can be controlled by firms using marketing variables such as price. As we will discuss in detail in Section 6, the presence of loyalty has economic implications for firms' pricing motives and equilibrium pricing outcomes. However, to make specific statements about how firms should set prices, we need to rule out that the structural state dependence effects in the data are due to some alternative form of consumer behavior. In this section, we consider the role of consumer search and product learning as possible alternative explanations. We do not postulate specific structural models of search or learning which would involve some strong structural assumptions on consumer behavior. Rather, we focus on aspects of consumer behavior that differentiate search or learning explanations from loyalty and that can be directly observed in our data.

□ **Search.** It is likely that consumers face search costs in the recall of the identities and the location of products in a store. Hoyer (1984) finds that consumers spent, on average, only 13 seconds "from the time they entered the aisle to complete their in-store decision." Furthermore, only 11% of consumers examined two or more products before making a choice in a given product category. Facing high search costs, consumers may purchase the products that they can easily recall or locate in the store. These products are likely to be the products which the consumer has purchased most recently. In this situation, consumers would display state dependence in product choice, as they may not be willing to pay the implicit search costs for investigating products other than those recently purchased.

FIGURE 8

DISTRIBUTION OF BRAND-SPECIFIC STATE DEPENDENCE COEFFICIENTS: REFRIGERATED ORANGE JUICE



The graph displays the pointwise posterior mean and 90% credibility region of the marginal density of the state dependence coefficient (γ^h), based on a five-component mixture-of-normals heterogeneity specification. We show the densities both for a model specification with a uniform (across-brands) state dependence coefficient and for a specification allowing for brand-specific state dependence coefficients (we show results for the four orange juice brands with the largest market shares).

In order to distinguish between state dependence due to loyalty and state dependence due to search costs, we exploit data on in-store advertising, sometimes termed *display advertising*. Retailers frequently add signs and even rearrange the products in the aisle to call attention to specific products. In the refrigerated orange juice category, 17.5% of the chosen items had an in-store display during the shopping trip (in the margarine category the incidence of displays is low).[11] A display can be thought of as an intervention that reduces a consumer's search cost.

In the marketing literature, it is sometimes assumed that consumers only choose among a subset of products in any given category, called the *consideration set*. Mehta, Rajiv, and Srinivasan (2003) construct a model for consideration set formation based on a fixed sample size search process. Using data for ketchup and laundry detergent products, they find that promotional activity, such as in-store displays, increases the likelihood that a product enters a consideration set. This work affirms the idea that in-store displays can reduce search costs.

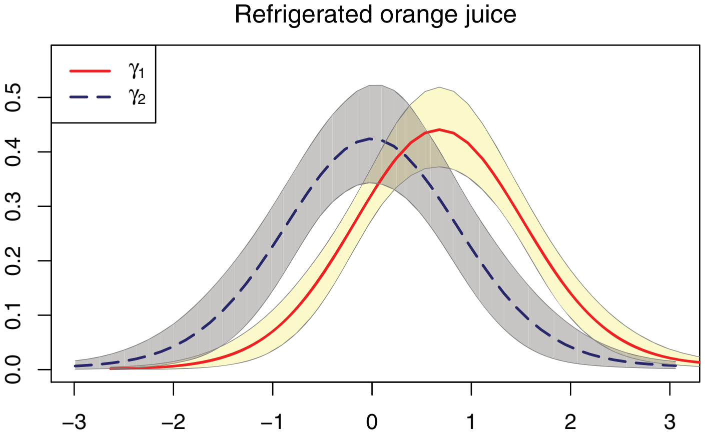
If displays affect demand via search costs, we should expect that a display increases the probability of a purchase. In addition, if a consumer has purchased a specific product in the past ($s_t = j$), then displays on *other* products should reduce the inertial effect or the tendency of the consumer to continue to purchase product j. This can be implemented by adding a specific interaction term to the baseline utility model:

$$\begin{aligned} u_{jt} &= \alpha_j + \eta p_{jt} + \gamma_1 \mathbb{I}\{s_t = j\} + \gamma_2 \mathbb{I}\{s_t \neq j\} \cdot \mathbb{I}\{display_{jt} = 1\} \\ &\quad + \lambda \mathbb{I}\{display_{jt} = 1\} + \epsilon_{jt}. \end{aligned} \tag{14}$$

[11] There is independent variation between displays and price discounts: no correlation between the display dummy variable and the level of prices exceeds 0.4 in absolute magnitude.

FIGURE 9

TESTING FOR SEARCH



The graph displays the pointwise posterior mean and 90% credibility region of the marginal density of the coefficients γ_1 and γ_2 in model (14). γ_1 is the main state dependence coefficient, and γ_2 measures the extent to which displays moderate the state dependence effect of past purchases. We expect that $\gamma_1 = \gamma_2$ if state dependence entirely proxies for search costs and if search costs disappear in the presence of a display. The results are based on a five-component mixture-of-normals heterogeneity specification. We only present results for refrigerated orange juice, as the incidence of displays is low in the margarine category.

To illustrate the coding of the interaction term in (14), consider the case of two brands and various display and purchase state conditions. If the consumer has purchased brand 1 in the past ($s_t = 1$) and neither brand is on display, then the utility for brand 1 relative to brand 2 is increased by γ_1. If brand 1 is on display, the utility difference increases by λ. If brand 2 is also on display, the main effect of display, λ, cancels out, and the interaction term turns on with the potential to offset the inertia effect. The difference between the utility for brand 1 and brand 2 due to state dependence and displays will be $\gamma_1 - \gamma_2$. Thus, γ_2 measures the extent to which displays moderate the state dependence effect of past purchases. If state dependence entirely proxies for search costs and if search costs disappear in the presence of a display, then we expect that $\gamma_1 = \gamma_2$.

Figure 9 plots the estimated marginal distributions of γ_1 and γ_2 (we only show results for orange juice, as we observe only few instances of displays in the margarine category). As before, the distribution of state dependence, γ_1, is centered at a positive value. However, the distribution of γ_2 is centered at zero. This result suggests that displays do not moderate the effect of past choices on current product purchases. We conclude that the measured state dependence is not merely a reduced-form effect that proxies for in-store search costs.

In spite of the lack of a moderating effect, the addition of a display main effect improves the model fit, increasing the log marginal likelihood from −4434 to −4360. Adding the interaction effect of displays and past purchase has a much smaller improvement on fit, increasing the log marginal likelihood from −4360 to −4339. Whatever the interpretation of the main effect of displays, it is unlikely that the estimated state dependence effect proxies for search behavior.

□ **Learning.** Consumers may have imperfect knowledge about the quality of products, in which case the consumption of a product may provide information about its true quality. Such learning about product quality may create inertia in choices over time. For example, suppose a consumer prefers brand B to brand A under perfect information. However, initially the consumer has only imperfect knowledge of the product's quality, and expects that the utility from consuming A is larger than the utility from consuming B. We then observe the consumer buying brand A until she gains experience with brand B, for example if she tries B when the product is on promotion.

If learning drives our state dependence findings, we would expect that experienced consumers in the category would exhibit a lower degree of state dependence than inexperienced consumers. In a model with learning, a consumer's choice process eventually converges to the predictions of a static choice model as product uncertainty is resolved. To proxy for shopping experience, we introduce a dummy for whether the primary shopper in the household is over 35 years old. Let θ^h be the vector of household parameters (including brand intercepts, price, and the inertia term). We partition θ^h into a part associated with the experienced shopper dummy and into residual unobserved heterogeneity that follows the mixture-of-normals distribution:

$$\begin{aligned} &\theta^h = \delta z^h + u^h, \\ &u^h \sim N(\mu_{ind}, \Sigma_{ind}), \quad ind \sim MN(\pi). \end{aligned} \tag{15}$$

δ is a vector which allows the means of all model coefficients to be altered by the experienced shopper dummy, z^h.

We find that the model fit decreases slightly by the addition of the experienced shopper dummy (Table 3). The element of δ that allows for the possibility of shifting the distribution of the state dependence coefficient is imprecisely estimated with a posterior credibility region that covers 0. For margarine, the posterior mean of this element is .17 with a 95% Bayesian credibility region of $(-.25, .60)$. For orange juice, the mean is .12 with a 95% Bayesian credibility region of $(-1.9, 1.75)$. We conclude that there is no evidence that the degree of state dependence differs across experienced and inexperienced shoppers.

A more powerful test of the learning hypothesis involves exploiting the fundamental difference between the loyalty and learning models in terms of the implications for the behavior of the choice process. If structural state dependence reflects loyalty, then as long as the exogenous variables (price, in our case) follow a stationary process, the choice process will also be stationary. However, in any model where learning is achieved through purchase and consumption, the choice process will be nonstationary. The consumers' posterior distributions of product quality will tighten as more consumption experience is obtained and consumers will exhibit a lower degree of state dependence over time. Eventually, consumers will behave in accordance with a standard choice model with no uncertainty.

We examine whether there is nonstationarity in the choice data, as would be implied by the learning model. Our panel is reasonably long and we expect that consumers will learn as they obtain more consumption experience with a brand. We define brand-level consumption experience as the cumulative number of purchases of the brand, E_{jt}. We can interact the state dependence variable with this new experience variable to provide a means of comparing the learning and loyalty models:

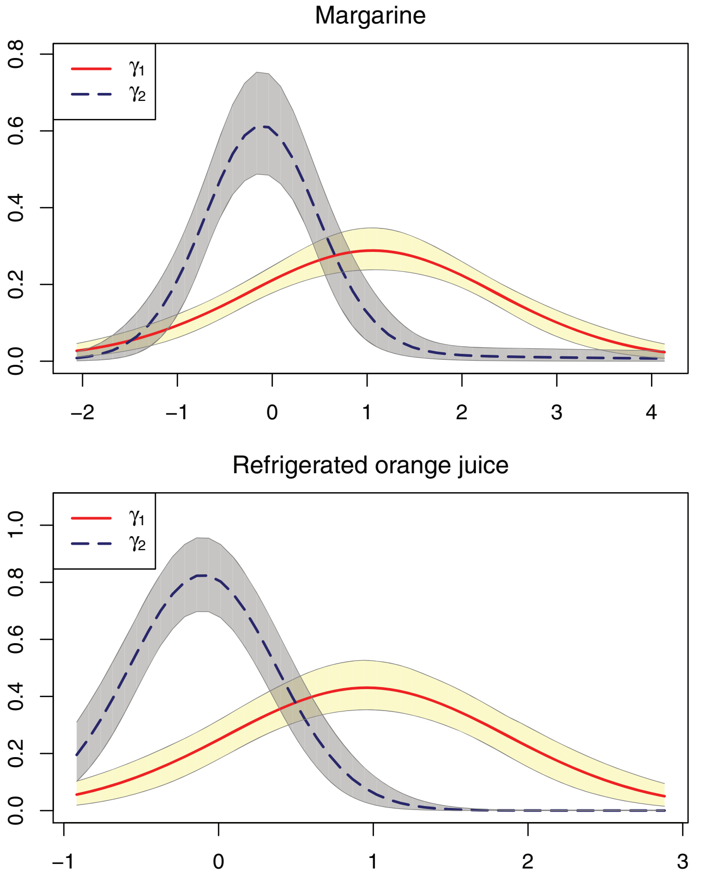
$$u_{jt} = \alpha_j + \eta_j p_{jt} + \gamma_1 \mathbb{I}\{s_t = j\} + \gamma_2 \mathbb{I}\{s_t = j\} \cdot E_{jt} + \lambda E_{jt} + \epsilon_{jt}. \tag{16}$$

As the experience variable adds additional information to the choice model, we should not directly compare the log marginal likelihood values of the interaction model (16) and the baseline model (1). The hypothesis that state dependence proxies for learning has implications for the interaction term in equation (16). Under learning, the interaction term should reduce state dependence as brand experience accumulates, that is, $\gamma_2 < 0$. Table 3 provides the log marginal likelihood values for a model with the interaction term, γ_2, and a model containing only a main effect of brand experience, γ_1. The marginal likelihood values increase by fairly small amounts when the interaction is added, 34 points in the margarine category and 4 points in the refrigerated orange juice category. Figure 10 verifies that the interaction terms are centered at 0 and contribute little to the model.

Of course, learning may only be relevant in situations where consumers have little consumption experience. Substantial evidence for learning has been found for new products by Ackerberg (2003) and Osborne (2007). Moshkin and Shachar (2002) find that learning explains

FIGURE 10

TESTING FOR LEARNING



The graph displays the pointwise posterior mean and 90% credibility region of the marginal density of the coefficients γ_1 and γ_2 in model (16). γ_1 is the main state dependence coefficient, and γ_2 represents the effect of the interaction between the purchase state and brand consumption experience, defined as the cumulative number of purchases of the brand. We expect that $\gamma_2 < 0$ if state dependence proxies for learning. The results are based on a five-component mixture-of-normals heterogeneity specification.

findings of state dependence for television programs, a product category with a very large and frequent number of new products. In our case, the same products have been in the marketplace for a considerable period of time. The households in the data might be expected to show little evidence of learning, given their experience with the brands prior to their involvement in the panel. This underscores the importance of a flexible model of heterogeneity. As a number of authors have noted, it is hard to distinguish learning models with heterogeneous initial priors from a standard choice model with brand preference heterogeneity. Indeed, Shin, Misra, and Horsky (2010) fit a learning model to a product category populated by well-established products. Once they supplement their data with survey data on household priors over product qualities, they measure very little learning.

6. The economic implications of state dependence

■ So far, we have established that there is robust evidence for structural state dependence in our data. Furthermore, the patterns of state dependence in the data are consistent with loyalty, a

TABLE 4 Dollar Value of Loyalty

Quantile (%)	Dollar Value	Dollar Value/Mean Price
	Margarine	
10	−0.11	−0.09
25	0.04	0.03
50	0.14	0.12
75	0.49	0.41
90	0.84	0.70
	Orange Juice	
10	0.12	0.04
25	0.27	0.10
50	0.56	0.21
75	1.15	0.42
90	2.09	0.77

form of state dependence whereby the utility from a product changes due to a past purchase or consumption experience, but not with search or learning. In this section, we explore the economic implications of loyalty.

□ **The dollar value of loyalty.** The inclusion of the outside option in the model enables us to assign money metric values to our model parameters by rescaling them by the price parameter (i.e., the marginal utility of income). The ratio $-\gamma/\eta$ represents the dollar equivalent of the utility premium induced by loyalty. Note that, even though there are no monetary costs associated with switching among brands, this ratio can be interpreted as a switching cost. As such, structural state dependence in the form of loyalty is a special case of switching costs (Klemperer, 1995; Farrell and Klemperer, 2007). We elaborate on the economic implications of this point in the next subsection. In this subsection, we focus on the actual dollar amounts of the switching costs.

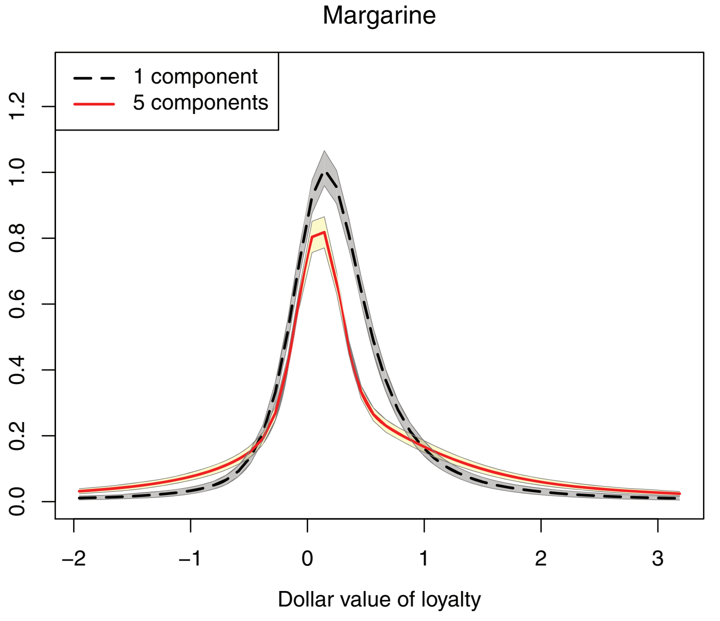
Table 4 displays selected quantiles from the distribution of the dollar loyalty premium across the population of households. Some of the values on which this distribution puts substantial mass are rather large, others are small. To provide some sense of the magnitudes of these values, we also compute the ratio of the dollar loyalty premium to the average price of the products. For margarine products, the median dollar value of loyalty is 12% of the average product price; for orange juice, the ratio is higher at 21%. However, there is a good deal of dispersion in the dollar loyalty value. At the 75th percentile of the distribution, the dollar loyalty value is 41% of the purchase price for margarine and 42% for orange juice. These are large values and of the order of many examples of standard economic (as opposed to psychologically derived) switching costs. For example, a cell phone termination penalty of $150 might be much less than total cell phone expenditures over the expected length of the contract. Another example of switching costs among packaged goods is razors and razor blades: a consumer needs to purchase a new razor when switching the type of razor blades. Here the monetary switching cost is small relative to razor blade prices (Hartmann and Nair, 2010).

Figure 11 illustrates the economic importance of controlling adequately for heterogeneity in the empirical estimation of structural state dependence in the form of loyalty. The five-component mixture-of-normals model generates a fitted density of the dollar value of loyalty or switching costs that is centered more closely to zero than the one-component model. This finding implies that the usual normal heterogeneity specification may overstate the degree of loyalty.

□ **The implications of structural state dependence for pricing.** An important component of the empirical analysis herein is the distinction between inertia as loyalty, a particular form of structural state dependence, versus inertia as unobserved heterogeneity or autocorrelated taste shocks. This distinction has both qualitative and quantitative implications for firms' pricing decisions on the supply side. The implications of brand choice inertia for firms' pricing decisions

FIGURE 11

DISTRIBUTION OF THE DOLLAR VALUE OF LOYALTY MARGARINE



The graph displays the pointwise posterior mean and 90% credibility region of the marginal density of the dollar value of loyalty, defined as $-\gamma^h/\eta^h$. The results are based on a five-component mixture-of-normals heterogeneity specification. For comparison purposes, we also show the results from a one-component heterogeneity specification.

and for equilibrium pricing outcomes differ depending on the source of the inertia. If inertia is due to autocorrelation in brand utilities or proxies for unobserved preference heterogeneity, firms cannot control the evolution of consumer preferences and, hence, there are no dynamic pricing incentives. In contrast, under loyalty, firms do face dynamic pricing incentives. Firms can use prices to influence current brand choices and, thus, influence future demand. Below, we compare the pricing incentives under each of these sources of inertia. We then conduct a simulation exercise to illustrate how, besides exhibiting qualitatively different pricing incentives, these alternative sources of inertia can lead to economically significant differences in equilibrium pricing outcomes.

To formalize the distinction in pricing incentives, consider a market with J firms competing in prices over time, $t = 0, 1, \dots$. The market is populated by a continuum of households characterized by the parameter vector $\theta \in \Theta$. $\phi(\theta)$ is the density of type θ households. Let $x_t(\theta) = (x_{1t}(\theta), \dots, x_{Jt}(\theta))$ denote the fraction of type θ households who are loyal to each of the J products. $\Pr\{j \mid \theta, p_t, s_t\}$ is the choice probability of household type θ for product j, given the price vector p_t and the loyalty state $s_t \in \{1, \dots, J\}$. Demand for product j is then given by

$$d_j(p_t, x_t) = \int_\Theta \left(\sum_{k=1}^{J} x_{kt}(\theta) \Pr\{j \mid \theta, p_t, k\} \right) \phi(\theta)\, d\theta, \tag{17}$$

where the mapping $x_t : \theta \to x_t(\theta)$ denotes the state of the market. The evolution of x_t over time can easily be derived from the household choice probabilities. In particular, $x_{j,t+1}(\theta)$, the fraction of type θ households loyal to product j in period $t+1$, is given by all type θ households who either bought j in period t or were already loyal to j in period t and chose the outside option. Conditional on p_t, the evolution of x_t is deterministic and can be denoted by $x_{t+1} = f(x_t, p_t)$.

Firms choose prices based on x_t, which contains all time-varying, payoff-relevant information about the market. Denote these pricing strategies by $\sigma_j : x \to p_j$. Conditional on

$\sigma_{-j} = (\sigma_1, \ldots, \sigma_{j-1}, \sigma_{j+1}, \ldots, \sigma_J)$, firm j's present value, given that it chooses a dynamically optimal pricing strategy, satisfies the Bellman equation

$$V_j(x_t) = \sup_{p_{jt}} \left\{ (p_{jt} - c_j) d_j(p_{jt}, \sigma_{-j}(x_t), x_t) + \beta V_j(x_{t+1}) \right\}, \tag{18}$$

where $x_{t+1} = f(x_t, p_{jt}, \sigma_{-j}(x_t))$. Here, c_j is the marginal cost of firm j and β is the discount factor.

The characterization of the pricing problem in equation (18) shows that structural state dependence in the form of loyalty gives rise to a nontrivial dynamic pricing problem. The elasticity of demand decreases in the number of loyal customers, and hence firms have an incentive to raise current prices if the current loyalty state increases. However, higher prices also affect the future state of the market, x_{t+1}. If firms lower their current price, x_{t+1} will increase and firms will thus face higher and less elastic demand in period $t+1$. This dynamic pricing problem is a special case of pricing under switching costs, and the two opposing incentives are typically called the *harvesting motive* and the *investment motive* in the switching cost literature (see the discussion in Klemperer, 1995 and Farrell and Klemperer, 2007). Dubé, Hitsch, and Rossi (2009) show that as the degree of state dependence increases, equilibrium prices either rise or fall depending on the relative strengths of the harvesting and investment motives.

In contrast, consider the pricing problem in the absence of loyalty. Household heterogeneity is still captured by the density $\phi(\theta)$. The heterogeneity allows consumers to have strong preferences for specific brands and, hence, to exhibit high repeat-purchase behavior. However, because utility is not affected by past product choices, demand is not a function of the loyalty states, x_t. Hence, current-period profits and the present value of each product or firm as described by the Bellman equation (18) do not depend on x_t. Therefore, the optimal prices can be found by maximizing static, per-period profits, and the optimal prices do not vary over time.

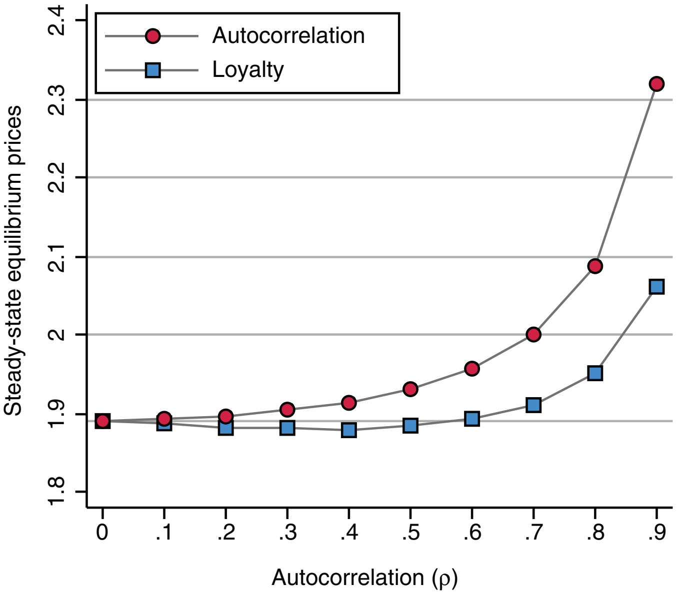
Dynamic pricing incentives are also absent if the random utility components are autocorrelated. For example, suppose that the latent utility of product j contains the component $\omega_{jt} = \rho\omega_{j,t-1} + \nu_{jt}$, where $\nu_{jt} \sim N(0, \sigma_\nu^2)$ and ρ captures the degree of autocorrelation. By assumption, ν_{jt} is independent of prices and i.i.d. across consumers and time. Therefore, the stationary distribution of the autocorrelated utility components across consumers is normal with mean 0 and variance $\sigma_\nu^2/(1-\rho^2)$. Market demand can be obtained by integrating over this distribution for each household type and, thus, from the firm's point of view, autocorrelation in utilities is simply another form of customer heterogeneity. The firms cannot control the distribution of the autocorrelated utility terms over time. Hence, as in the case of preference heterogeneity discussed above, the optimal prices maximize static profits and are time invariant.

We now present an example that shows how, in addition to exhibiting different pricing incentives, the two sources of inertia can generate economically significant differences in equilibrium pricing outcomes. Suppose an analyst observes consumer choice data in a market with two symmetric firms, estimates household preferences based on the data, and predicts equilibrium prices based on the demand estimates and cost information. Suppose also that consumer choices exhibit inertia due to autocorrelated taste shocks but no loyalty. The analyst, however, makes the false assumption that the observed inertia in choices is entirely due to loyalty and hence estimates household preferences using a model with a state dependence term. We compare the analyst's prediction of equilibrium prices with the prices corresponding to the true model with autocorrelated errors. We obtain the equilibrium prices from a numerical solution of a Markov perfect equilibrium; see Dubé, Hitsch, and Rossi (2009) for details. We conduct this comparison for several different degrees of inertia in the data as given by the autocorrelation parameter ρ.[12] Figure 12 compares the true equilibrium prices and the analyst's predictions for

[12] The household choice data are generated using the parameter values $\alpha_1 = \alpha_2 = 1$, $\eta = -1$, and $\sigma_\nu^2 = 1$. σ_ν^2 is known to the analyst, which corresponds to a scale normalization of the random utility terms. The firms' unit cost of production is $c = 1$ and the discount factor is $\beta = 0.998$.

FIGURE 12

EQUILIBRIUM PRICES UNDER STATE DEPENDENCE AND AUTOCORRELATION



The graph displays the (symmetric) steady-state equilibrium prices from a model with autocorrelated random utility terms, and contrasts these "true" prices to the price predictions if the inertia in the brand choice data were attributed to structural state dependence in the form of loyalty.

the different values of ρ. The loyalty model was estimated using a large data set[13] such that we can ignore the tiny amount of parameter uncertainty in the price predictions. As discussed above, higher levels of ρ are analogous to an increased dispersion of consumer brand preferences. This increase in preference heterogeneity softens price competition, and the true equilibrium prices rise monotonically in ρ. As ρ increases, the analyst's estimate of the coefficient on the loyalty state increases. The analyst believes that inertia is generated by loyalty and, therefore, that firms can control the future loyalty states. The downward pressure exerted through the investment motive causes the predicted equilibrium prices under the incorrect loyalty model to be smaller than the true equilibrium prices under autocorrelation. For small values of ρ, the predicted equilibrium prices under loyalty even fall relative to the case of no inertia. The difference in predicted prices is most pronounced for large values of ρ. For example, if $\rho = 0.9$, the true prices are 12.4% larger than the analyst's prediction. In summary, incorrectly specifying the source of inertia can lead to quantitatively different predictions for equilibrium pricing behavior.

7. Conclusions

■ We find strong evidence that observed inertia in consumer choices for margarine and refrigerated orange juice is driven by structural state dependence, even after controlling for various forms of spurious state dependence. In particular, our findings of structural state dependence are robust to a semiparametric mixture-of-normals specification for time-invariant preference heterogeneity. Our findings of structural state dependence are also robust to a test for autocorrelated taste shocks.

Unlike much of the previous empirical work, we explore the underlying source of inertia by comparing three potential economic explanations: loyalty, consumer search, and learning. The

[13] The simulated data contain 2000 shopping trips for 2000 households.

structural interpretation of the loyalty model is that when a specific brand is purchased, it is accorded a utility premium on future choice occasions, much like a switching cost. In search models, consumers may persist in purchasing one brand if the costs of exploring other options are high. In learning models, what appears to be inertia can arise because of imperfect information about product quality. Products which a consumer has previously consumed have less uncertainty in quality evaluation, and this may make consumers reluctant to switch to alternative products for which there is greater quality uncertainty. We find that the form of inertia in our data is consistent with loyalty, but not with search or learning.

In our model specification, we assume that consumers are myopic and do not consider the impact of current purchase decisions on future utility. We think of state dependence in the form of loyalty as a subconscious (or psychological) switching cost and we do not expect consumers to choose among brands in anticipation of future loyalty states. In contrast, other work in the empirical literature on consumer choice has considered forward-looking behavior in the presence of switching costs (e.g., Osborne, 2007) as well as in contexts such as stock-piling (e.g., Erdem, Imai, and Keane, 2003; Hendel and Nevo, 2006) or learning (e.g., Erdem and Keane, 1996).

Finally, we explore the economic implications of loyalty as the driving force of consumer inertia. First, we show that the magnitude of the underlying switching costs is economically large compared to typical shelf prices for the brands studied. Second, we show that the implications for pricing decisions and equilibrium pricing outcomes are markedly different when inertia arises due to loyalty as opposed to unobserved heterogeneity or autocorrelated errors. Therefore, the empirical distinction between structural state dependence and spurious state dependence is important for policy analysis. In companion pieces, Dubé, Hitsch, Rossi, and Vitorino (2008) and Dubé, Hitsch, and Rossi (2009), we explore the implications of estimated switching costs from our loyalty model for dynamic pricing under multiproduct monopoly and dynamic oligopoly, respectively.

Appendix A

□ **MCMC and prior settings.** The MCMC method applied here is a hybrid method with a customized Metropolis chain for the draw of the household-level parameters coupled with a standard Gibbs sampler for a mixture of normals conditional on the draws of household-level parameters. That is, once the collection of household parameters is drawn, the MCMC algorithm treats these as “data” and conducts Bayesian inference for a mixture of normals. Thus, there are “two” stages in the algorithm:

$$\theta_h \mid y_h, X_h, ind_h, \mu_{ind_h}, \Sigma_{ind_h} \quad h = 1, \ldots, H \tag{A1}$$

$$ind, \pi, \{\mu_k, \Sigma_k\} \mid \Theta. \tag{A2}$$

Θ is a matrix consisting of H rows, each with the θ_h parameters for each household, y_h is the vector choice observations for household h, and X_h is the matrix of covariates. The first stage of the MCMC in (A1) is a set of H Metropolis algorithms tuned to each household multinomial logit (MNL) likelihood. The tuning is done automatically without any “presampling” of draws and is done on the basis of a fractional likelihood that combines the household likelihood fractionally with the pooled MNL likelihood (for further details, see Rossi, Allenby, and McCulloch, 2005). It should be noted that this tuning is just for the Metropolis proposal distribution. This procedure avoids the problem of undefined likelihoods for tuning purposes. The household likelihood used in the posterior computations is not altered.

The second stage (A2) is a standard unconstrained Gibbs sampler for a mixture of normals. The “label-switching” problem for identification in a mixture of normals is not present in our application, as we are interested in the posterior distribution of a quantity which is label invariant, that is, the mixture-of-normals density itself. The priors used are

$$\pi \sim Dirichlet(a)$$
$$\mu_k \mid \Sigma_k \sim N\left(\overline{\mu}, a_\mu^{-1}\Sigma_k\right)$$
$$\Sigma_k \sim IW(\nu, \nu I).$$

The prior hyperparameters were assessed to provide proper but diffuse distributions. $a = (.5/K, K),\ a_{mu} = 1/16,\ \nu = dim(\theta_h) + 3$. The Dirichlet prior on π warrants further comment. The Dirichlet distribution is conjugate to the multinomial. $\sum a$ can be interpreted as the size of a prior sample of data for which the classification of θ_h “observations” is known. The number of observations of each “type” or mixture component is given by the appropriate element of a. Our prior says that each type is equally likely and that there is only a very small amount of information in the prior equal to a sample "size" of .5. As the number of normal components increases, we do not want to change how informative the prior is; this is why we scale the elements of the a vector by K.

Our computer code for this model can be found in the contributed R package, `bayesm`, available on the CRAN network of mirror sites (see function `rhierMnlRwMixture`) (http://cran.r-project.org/).

Appendix B

□ **The conditional expectation of ϵ_j given that product j is the optimal choice.** Let $u_j(x)$ be the nonrandom component of the latent utility from product j, which depends on the variables in x. A consumer chooses product j among all J options if and only if

$$\epsilon_j \geq u_k(x) - u_j(x) + \epsilon_k \quad \text{for all } k \neq j.$$

Let $\epsilon_{-j} \equiv (\epsilon_1, \ldots, \epsilon_{j-1}, \epsilon_{j+1}, \ldots, \epsilon_J)$ and define

$$\psi_j(x, \epsilon_{-j}) \equiv \max_{k \neq j}\{u_k(x) - u_j(x) + \epsilon_k\}.$$

Then j is the optimal choice (denoted by d) if and only if $\epsilon_j \geq \psi_j(x, \epsilon_{-j})$. We assume that all ϵ_js are independent with a density p_j that is positive everywhere on $\mathbb{R}^J$. Therefore, the conditional density of ϵ_j is

$$p_j(\epsilon_j \mid d = j, \epsilon_{-j}, x) = \frac{p_j(\epsilon_j)}{1 - \int_{-\infty}^{\psi_j(x,\epsilon_{-j})} p_j(\epsilon)\, d\epsilon}$$

for all $\epsilon_j \geq \psi_j(x, \epsilon_{-j})$, and $p(\epsilon_j \mid d = j, \epsilon_{-j}, x) = 0$ otherwise.

Consider two vectors x_0 and x_1 such that $\omega_0 \equiv \psi_j(x_0, \epsilon_{-j}) < \psi_j(x_1, \epsilon_{-j}) \equiv \omega_1$. Define $\pi \equiv \Pr\{\epsilon_j < \omega_1 \mid d = j, \epsilon_{-j}, x_0\}$. Then

$$\begin{aligned}
\mathbb{E}[\epsilon_j \mid d = j, \epsilon_{-j}, x_0] &= \pi \mathbb{E}[\epsilon_j \mid \epsilon_j < \omega_1, d = j, \epsilon_{-j}, x_0] + (1 - \pi)\mathbb{E}[\epsilon_j \mid \epsilon_j \geq \omega_1, d = j, \epsilon_{-j}, x_0] \\
&< \pi\omega_1 + (1 - \pi)\mathbb{E}[\epsilon_j \mid \epsilon_j \geq \omega_1, d = j, \epsilon_{-j}, x_0] \\
&= \pi\omega_1 + (1 - \pi)\mathbb{E}[\epsilon_j \mid d = j, \epsilon_{-j}, x_1] \\
&< \mathbb{E}\left[\epsilon_j \mid d = j, \epsilon_{-j}, x_1\right].
\end{aligned}$$

Both inequalities are strict because $p_j(\epsilon_j \mid d = j, \epsilon_{-j}, x_0) > 0$ for $\epsilon_j \in [\omega_0, \omega_1)$. By the law of iterated expectations,

$$\mathbb{E}[\epsilon_j \mid d = j, x_0] = \mathbb{E}_{\epsilon_{-j}}[\mathbb{E}[\epsilon_j \mid d = j, \epsilon_{-j}, x_0]] < \mathbb{E}_{\epsilon_{-j}}[\mathbb{E}[\epsilon_j \mid d = j, \epsilon_{-j}, x_1]] = \mathbb{E}[\epsilon_j \mid d = j, x_1].$$

Suppose x_0 and x_1 represent two marketing environments that are identical apart from the price of product j, which is discounted in x_0 but not in x_1. If the discounted price leads to a strictly higher utility from product j, then $\psi_j(x_0, \epsilon_{-j}) < \psi_j(x_1, \epsilon_{-j})$, and hence $\mathbb{E}[\epsilon_j \mid d = j, x_0] = \mathbb{E}[\epsilon_j \mid d = j, x_1]$.

References

ACKERBERG, D.A. "Advertising, Learning, and Consumer Choice in Experience Good Markets: An Empirical Examination." *International Economic Review,* Vol. 44 (2003), pp. 1007–1040.

ALLENBY, G.M. AND ROSSI, P.E. "Marketing Models of Consumer Heterogeneity." *Journal of Econometrics,* Vol. 89 (1999), pp. 57–78.

CHAMBERLAIN, G. "Heterogeneity, Omitted Variable Bias, and Duration Dependence." In J.J. Heckman and B. Singer, eds., *Longitudinal Analysis of Labor Market Data*. Cambridge, UK: Cambridge University Press, 1985.

DAWID, A.P. "Prequential Analysis, Stochastic Complexity and Bayesian Inference." In J.M. Bernardo, J.O. Berger, and A.F.M. Smith, eds., *Bayesian Statistics 4*. Oxford: Oxford University Press, 1992.

DUBÉ, J.-P., HITSCH, G.J., ROSSI, P.E., AND VITORINO, M.A. "Category Pricing with State-Dependent Utility." *Marketing Science,* Vol. 27 (2008), pp. 417–429.

——, ——, AND ——. "Do Switching Costs Make Markets Less Competitive?" *Journal of Marketing Research,* Vol. 46 (2009), pp. 435–445.

ERDEM, T. AND KEANE, M.P. "Decision-Making under Uncertainty: Capturing Dynamic Brand Choice Processes in Turbulent Consumer Goods Markets." *Marketing Science,* Vol. 15 (1996), pp. 1–20.

—— AND SUN, B. "Testing for Choice Dynamics in Panel Data." *Journal of Business and Economic Statistics,* Vol. 19 (2001), pp. 142–152.

——, IMAI, S., AND KEANE, M.P. "Brand and Quantity Choice Dynamics under Price Uncertainty." *Quantitative Marketing and Economics,* Vol. 1 (2003), pp. 5–64.

ESCOBAR, M.D. AND WEST, M. "Bayesian Density Estimation and Inference Using Mixtures." *Journal of the American Statistical Association,* Vol. 90 (1995), pp. 577–588.

FARRELL, J. AND KLEMPERER, P. "Coordination and Lock-In: Competition with Switching Costs and Network Effects." In M. Armstrong and R. Porter, eds., *Handbook of Industrial Organization,* Vol. 3. Elsevier B.V., 2007.

FRANK, R.E. "Brand Choice as a Probability Process." *Journal of Business,* Vol. 35 (1962), pp. 43–56.

HARTMANN, W.R. AND NAIR, H.S. "Retail Competition and the Dynamics of Demand for Tied Goods." *Marketing Science,* Vol. 29 (2010), pp. 366–386.

HECKMAN, J.J. "Statistical Models for Discrete Panel Data." In C.F. Manski and D. McFadden, eds., *Structural Analysis of Discrete Data with Econometric Applications*. Cambridge, Mass.: MIT Press, 1981.

HENDEL, I. AND NEVO, A. "Measuring the Implications of Sales and Consumer Inventory Behavior." *Econometrica,* Vol. 74 (2006), pp. 1637–1673.

HOYER, W.D. "An Examination of Consumer Decision Making for a Common Repeat Purchase Product." *Journal of Consumer Research,* Vol. 11 (1984), pp. 822–829.

KEANE, M.P. "Modeling Heterogeneity and State Dependence in Consumer Choice Behavior." *Journal of Business and Economic Statistics,* Vol. 15 (1997), pp. 310–327.

KLEMPERER, P. "Competition When Consumers Have Switching Costs: An Overview with Applications to Industrial Organization, Macroeconomics, and International Trade." *Review of Economic Studies,* Vol. 62 (1995), pp. 515–539.

LOPES, H. AND GAMERMAN, D. *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. London: Chapman Hill, 2006.

MASSY, W.F. "Order and Homogeneity of Family Specific Brand-Switching Processes." *Journal of Marketing Research,* Vol. 3 (1966), pp. 48–54.

MEHTA, N., RAJIV, S., AND SRINIVASAN, K. "Price Uncertainty and Consumer Search: A Structural Model of Consideration Set Formation." *Marketing Science,* Vol. 22 (2003), pp. 58–84.

MOSHKIN, N.V. AND SHACHAR, R. "The Asymmetric Information Model of State Dependence." *Marketing Science,* Vol. 21 (2002), pp. 435–454.

NEWTON, M.A. AND RAFTERY, A.E. "Approximate Bayesian Inference with the Weighted Likelihood Bootstrap." *Journal of the Royal Statistical Society, Series B,* Vol. 56 (1994), pp. 3–48.

OSBORNE, M. "Consumer Learning, Switching Costs, and Heterogeneity: A Structural Examination." Discussion Paper 07–10, Economic Analysis Group, Department of Justice, 2007.

ROSSI, P.E., MCCULLOCH, R.E., AND ALLENBY, G.M. "The Value of Purchase History Data in Target Marketing." *Marketing Science,* Vol. 15 (1996), pp. 321–340.

——, ALLENBY, G.M., AND MCCULLOCH, R.E. *Bayesian Statistics and Marketing*. New York: John Wiley & Sons, 2005.

SEETHARAMAN, P.B. "Modeling Multiple Sources of State Dependence in Random Utility Models: A Distributed Lag Approach." *Marketing Science,* Vol. 23 (2004), pp. 263–271.

——, AINSLIE, A., AND CHINTAGUNTA, P.K. "Investigating Household State Dependence Effects across Categories." *Marketing Science,* Vol. 36 (1999), pp. 488–500.

SHIN, S., MISRA, S., AND HORSKY, D. "Disentangling Preferences and Learning in Brand Choice Models." *Marketing Science* (2010, forthcoming).

SHUM, M. "Does Advertising Overcome Brand Loyalty? Evidence from the Breakfast-Cereals Market." *Journal of Economics and Management Strategy,* Vol. 13 (2004), pp. 241–272.