NBER WORKING PAPER SERIES

STATE-DEPENDENT DEMAND ESTIMATION WITH INITIAL CONDITIONS CORRECTION

Andrey Simonov
Jean-Pierre H. Dubé
Günter J. Hitsch
Peter E. Rossi

Working Paper 26217
http://www.nber.org/papers/w26217

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 2019

Dube and Hitsch are grateful to the Kilts Center for Marketing for research support. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

State-Dependent Demand Estimation with Initial Conditions Correction
Andrey Simonov, Jean-Pierre H. Dubé, Günter J. Hitsch, and Peter E. Rossi
NBER Working Paper No. 26217
September 2019
JEL No. D12,L66,M3

ABSTRACT

We analyze the initial conditions bias in the estimation of brand choice models with structural state dependence. Using a combination of Monte Carlo simulations and empirical case studies of shopping panels, we show that popular, simple solutions that mis-specify the initial conditions are likely to lead to bias even in relatively long panel datasets. The magnitude of the bias in the state dependence parameter can be as large as a factor of 2 to 2.5. We propose a solution to the initial conditions problem that samples the initial states as auxiliary variables in an MCMC procedure. The approach assumes that the joint distribution of prices and consumer choices, and hence the distribution of initial states, is in equilibrium. This assumption is plausible for the mature consumer packaged goods products used in this and the majority of prior empirical applications. In Monte Carlo simulations, we show that the approach recovers the true parameter values even in relatively short panels. Finally, we propose a diagnostic tool that uses common, biased approaches to bound the values of the state dependence and construct a computationally light test for state dependence.

Andrey Simonov
Columbia Business School
90 Morningside Dr Apt 6C
New York, NY 10027
andsimonov@gmail.com

Jean-Pierre H. Dubé
University of Chicago
Booth School of Business 5807
South Woodlawn Avenue
Chicago, IL 60637
and NBER
jdube@chicagobooth.edu

Günter J. Hitsch
University of Chicago
Booth School of Business
5807 South Woodlawn Avenue
Chicago, IL 60637
guenter.hitsch@ChicagoBooth.edu

Peter E. Rossi
UCLA/Anderson School
110 Westwood Plaza B4.12
Los Angeles, CA 90095
peter.rossi@anderson.ucla.edu

1 Introduction

Testing for and measuring *brand loyalty*, or persistence in brand choice, remains one of the oldest and most central themes in the empirical quantitative marketing literature [e.g., Brown, 1952, Frank, 1962]. Typically, a consumer's complete, lifetime brand choice history is not observed, leading to an *initial conditions problem* [Heckman, 1981a]. The initial conditions problem arises when the initial observed loyalty state depends on unobserved aspects of consumer preferences, creating a specification error in approaches that treat the initial state as exogenous.

Due to the computational complexity of modeling the dependence of the initial conditions on unobserved consumer tastes, the extant literature has typically relied on assumptions that ignore the initial conditions problem. One approach assumes that the initial loyalty state is independent of preferences (initial state is exogenous) [e.g., Keane, 1997, Dubé et al., 2010], while the second one treats the first observed purchase occasion as having no loyalty (initial state is zero) [e.g., Seetharaman et al., 1999, Shum, 2004]. Both approaches lead to inconsistent estimates of loyalty for fixed T and exhibit a small-sample bias (in T) in the estimates. While this bias dissipates as $T \to \infty$, the order of T required to eliminate the bias in practice can potentially be infeasible even with long panels spanning many years of shopping history.

We investigate the small-sample biases associated with ignoring the initial conditions problem when estimating a state-dependent brand choice model. In a series of Monte Carlo simulations, we show that the biases due to misspecified initial conditions is substantial even for panels that are longer than conventional CPG (consumer packaged goods) shopping databases, like those collected by Nielsen and IRI. We further show that the procedures used in the extant literature exhibit systematic forms of bias: in short panels, assuming an exogenous initial state leads to an upward bias in the estimated degree of loyalty, while assuming no initial loyalty (setting initial state to zero) leads to attenuation bias. These two biased approaches can be used to construct computationally simple estimators that approximately bound the degree of state dependence as a preliminary test that does not require modeling the initial conditions.

We then propose a Bayesian estimator that incorporates the endogenous initial conditions into the likelihood function. This approach is based on the assumption that the stochastic process governing the joint distribution of prices and consumer choices is in equilibrium. This assumption is plausible in the vast majority of extant empirical applications, which study state dependence in *mature* consumer packaged goods (CPG) markets. To solve the computational challenge of evaluating the likelihood, we propose a simulation procedure that treats the initial state as an auxiliary variable and yields draws from the joint distribution of consumer preference parameters and initial states as part of the Gibbs sampling procedure. To sample the initial states, we compute the conditional probability of being in each state given the current draws of consumer preferences. The sampler also accounts for the dynamics in prices, which we treat as the outcome of an estimated Markov process using the observed prices.

We test the performance of the proposed estimator in a series of Monte Carlo simulations and compare it to approaches that treat the initial conditions as exogenous. We find that our proposed approach recovers the true values of brand loyalty quite well, even in panels of moderate length. Interestingly, the estimator that also addresses the dynamics in prices only performs slightly better than the estimator that treats prices as i.i.d., even when the true data-generating process involves Markovian prices. In contrast, naive estimators that ignore the initial conditions problem generate biases, even in long panels. Using the two naive approaches as bounds on the true value of the state dependence, our Monte Carlo simulations show that the bounds converge to the true value as we increase the panel length. We also find that the two bounds coincide at zero for a data-generating process with no state dependence.

We compare the same set of estimators in two empirical case studies of the margarine category. The first case study comprises the 2.5-year panel shopping panel from Dubé et al. [2010]. The second case study comprises a 6-year panel from the Nielsen-Kilts Homescan (HMS) data. The estimates confirm the conclusions from the Monte Carlo simulations. As explained above, we can use the two naive estimators to bound the degree of state dependence above zero. We therefore conclude that there is indeed state dependence in consumers' brand choices. Our proposed estimator accounts for the dependence of the initial conditions on unobserved consumer preferences and generates an estimate of the state dependence that lies between the two naive approaches, as expected. The magnitude of bias when ignoring the initial conditions problem is of a factor of 2-2.5. For instance, our results suggest that Dubé et al. [2010] under-estimated the degree of brand loyalty by more than half.

These findings add to a large literature exploring biases in empirical models of state-dependent brand choice. The initial conditions bias we investigate herein builds on the seminal work of Heckman [1981a] and is similar to the biases associated with incorrect assumptions about initial beliefs in learning models [e.g., Shin et al., 2012], a related form of structural state dependence, and the initial inventory state variable in the literature on stock-piling and dynamic discrete choice [e.g., Erdem et al., 2003, Hendel and Nevo, 2006].[1] Our findings also build on the literature that has discussed the potential biases associated with unobserved heterogeneity in state-dependence models of brand choice [e.g., Heckman, 1991, Keane, 1997, Dubé et al., 2010]. However, controlling for unobserved heterogeneity does not resolve the initial conditions problem. Finally, our findings have supply-side implications for brand pricing. In particular, incorrectly assuming exogeneity in the initial observed loyalty state will bias the degree of brand loyalty, which in turn biases the dynamic incentives to firms when setting their prices [e.g., Dubé et al., 2008, 2009].

The rest of the paper is organized as follows. Section 2 formalizes the initial conditions' problem in a model of consumer discrete-choice demand with heterogeneity and state dependence, and proposes a correction procedure. Section 3 tests the correction procedure and "naive" ways to treat

[1] Heckman [1981a] considers a simple case of intercept heterogeneity and does not provide a computationally feasible solution to the problem.

the initial state in a Monte Carlo simulation exercise. Section 4 applies the estimation routines to two real-world datasets. Section 5 concludes.

2 Model and Estimation

2.1 Model

In this section, we focus on the familiar first-order Markov multinomial discrete choice model of demand [e.g., Chintagunta, 1998, Keane, 1997, Seetharaman et al., 1999, Dubé et al., 2010], which extends the discussion of binary choices in Heckman [1981b]. Consumers indexed by $h = 1, ..., H$ make discrete purchase decisions j from a choice set $\mathcal{A} = \{0, 1, \ldots, J\}$ during each of the time periods $t = 1, ..., T$. The $j = 0$ alternative indicates non-purchase.

The choice-specific utility of consumer h is:

$$u_j(s_t^h, \boldsymbol{\epsilon}_t^h; \boldsymbol{\Theta}^h) = \begin{cases} \beta_j^h + \gamma^h \mathbb{1}\{s_t^h = j\} + \epsilon_{jt}^h & \text{if } j \neq 0, \\ \epsilon_{0t}^h & \text{if } j = 0. \end{cases}$$

$\boldsymbol{\Theta}^h = \left(\beta_1^h, ..., \beta_J^h, \gamma^h\right)'$ is a vector that includes the consumer's taste parameters and $\epsilon_{jt}^h \sim i.i.d.$ EV$(0, 1)$ are random utilities. The state variable $s_t^h \in \{1, \ldots, J\}$ indicates the consumer's current loyalty state. When choosing the alternative that the consumer is loyal to, $j = s_t^h$, the consumer receives the utility component γ^h. Adding additional covariates in the utility function is straightforward, but for now we only consider the most minimal model structure to keep the exposition simple.

The consumer makes the discrete choice $y_t^h = j$ if and only if

$$u_j(s_t^h, \boldsymbol{\epsilon}_t^h; \boldsymbol{\Theta}^h) \geq u_k(s_t^h, \boldsymbol{\epsilon}_t^h; \boldsymbol{\Theta}^h) \qquad \forall k \neq j.$$

The corresponding probability that consumer h chooses alternative j is:

$$\Pr\left\{y_t^h = j | s_t^h, \boldsymbol{\Theta}^h\right\} = \frac{\exp\left(\beta_j^h + \gamma^h \mathbb{1}\{s_t^h = j\}\right)}{1 + \sum_{k=1}^J \exp\left(\beta_k^h + \gamma^h \mathbb{1}\{s_t^h = k\}\right)}. \tag{1}$$

In the model, the choice in the previous period determines the current period's loyalty state. If the consumer chooses $j \neq 0$, the loyalty state in the next period will be $s_{t+1}^h = j$. If the consumer chooses the outside option, $j = 0$, the loyalty state will remain unchanged, $s_{t+1}^h = s_t^h$. Therefore, the loyalty state follows a Markov Process where the transition probabilities are derived from the conditional choice probabilities. The probability that the loyalty state transitions from $s_t^h = j$ to

$s_{t+1}^h = k \neq 0$ is

$$\Pr\left\{s_{t+1}^h = k | s_t^h = j, \mathbf{\Theta}^h\right\} = \begin{cases} \Pr\left\{y_t^h = k | s_t^h, \mathbf{\Theta}^h\right\} & \text{if } k \neq j, \\ \Pr\left\{y_t^h = j | s_t^h, \mathbf{\Theta}^h\right\} + \Pr\left\{y_t^h = 0 | s_t^h, \mathbf{\Theta}^h\right\} & \text{if } k = j. \end{cases} \tag{2}$$

Following most of the prior literature, we assume that a consumer will never become loyal to the outside option. However, this is not an essential assumption and could be relaxed. For example, we could allow for a loyalty (or rather: non-loyalty) state $s_t^h = 0$, such that a consumer who is in this state does not receive the utility component γ^h for any choice.

The marginal probability of an observed choice in a given period depends on the probabilities of all possible choice histories up until that point. Formally, the marginal probability that consumer h chooses j in period t is

$$\Pr\{y_t^h = j | \mathbf{\Theta}^h\} = \sum_{\boldsymbol{s}_t^h \in \mathbb{S}_t} \Pr\{j | s_t^h, \mathbf{\Theta}^h\} \cdot \Pr\{s_t^h | s_{t-1}^h, \mathbf{\Theta}^h\} \cdots \Pr\{s_2^h | s_1^h, \mathbf{\Theta}^h\} \cdot \Pr\{s_1^h | \mathbf{\Theta}^h\}, \tag{3}$$

where $\boldsymbol{s}_t^h = (s_1^h, \ldots, s_t^h)$ and $\mathbb{S}_t$ is the t-fold Cartesian product of the set of loyalty states, $\{1, \ldots, J\}$. Furthermore, because s_t^h depends on s_{t-1}^h and the choice y_{t-1}^h, it is evident that the marginal probability of a loyalty state, $\Pr\{s_t^h | \mathbf{\Theta}^h\}$, depends on the entire prior choice process, and thus on $\mathbf{\Theta}^h$. Even if the data included the very first period when a consumer made a choice and the initial loyalty state, s_1^h and $\mathbf{\Theta}^h$ would in general be dependent. For example, adjusting for prices and other covariates, a consumer will be more likely to be loyal to her most preferred alternative, i.e. the alternative with the largest value of β_j^h. Only under some special conditions, such as when consumers are not loyal to any alternative in the initial period ($\Pr\{s_1^h = 0 | \mathbf{\Theta}^h\} = 1$ for all $\mathbf{\Theta}^h$), s_1^h will be independent of $\mathbf{\Theta}^h$.

We now consider the problem of statistical inference for the distribution of preference parameters. $F_{\mathbf{\Theta}}\left(\mathbf{\Theta}^h | \boldsymbol{\theta}\right)$ is the prior for the population distribution of the individual preference parameters with a vector of hyper-parameters $\boldsymbol{\theta}$. For each consumer, we observe the choices $\boldsymbol{y}^h = \left(y_1^h, \ldots, y_T^h\right)$.[2] $\ell(\boldsymbol{y}^h | \mathbf{\Theta}^h)$ is the multinomial logit likelihood-contribution for consumer h:

$$\ell(\boldsymbol{y}^h | \mathbf{\Theta}^h) = \sum_{j=1}^{J} \left(\prod_{t=1}^{T} \Pr\left\{y_t^h | s_t^h, \mathbf{\Theta}^h\right\} \cdot \Pr\{s_1^h = j | \mathbf{\Theta}^h\} \right). \tag{4}$$

Note the dependence of the likelihood on the distribution of the initial condition, $\Pr\left\{s_1^h = i | \mathbf{\Theta}^h\right\}$, which in general is dependent on the consumer-specific parameters $\mathbf{\Theta}^h$.

One approach to solve the initial conditions problem is to assume that the process generated by the sequence of choices is in equilibrium. In our model, all choice probabilities (1) are strictly positive, and hence the distribution of the loyalty state converges to the stationary distribution $\boldsymbol{\pi}(\mathbf{\Theta}^h) =$

[2] Assuming the same number of observations per consumer for notational convenience only.

$(\pi_1(\mathbf{\Theta}^h), \ldots, \pi_J(\mathbf{\Theta}^h))$. The probability of the initial condition is $\Pr\{s_1^h = j|\mathbf{\Theta}^h\} = \pi_j(\mathbf{\Theta}^h)$.

The evaluation of the initial condition probability becomes more complicated when we include covariates in the choice model. Typically, we will includes prices in the choice-specific indirect utility function:

$$u_j(\boldsymbol{p}_t, s_t^h, \boldsymbol{\epsilon}_t^h; \mathbf{\Theta}^h) = \begin{cases} \beta_j^h - \alpha^h p_{jt} + \gamma^h \mathbb{1}\{s_t^h = j\} + \epsilon_{jt}^h & \text{if } j \neq 0, \\ \epsilon_{0t}^h & \text{if } j = 0. \end{cases}$$

Assume that prices follow a Markov process with transition probability $F_{\boldsymbol{p}}(\boldsymbol{p}_{t+1}|\boldsymbol{p}_t)$, and let $F_{\boldsymbol{p}_1}(\boldsymbol{p}_1)$ be the distribution of prices in the initial period. Let $G_t(\tilde{\boldsymbol{p}})$ be the implied distribution over the sequence of price vectors, $\tilde{\boldsymbol{p}} = (\boldsymbol{p}_1, \ldots, \boldsymbol{p}_t)$. Now the marginal probability that consumer h chooses j in period t is:

$$\Pr\{y_t^h = j|\mathbf{\Theta}^h\} = \int \sum_{\boldsymbol{s}_t^h \in \mathbb{S}_t} \Pr\{j|\boldsymbol{p}_t, s_t^h, \mathbf{\Theta}^h\} \cdot \prod_{\tau=2}^{t} \Pr\{s_\tau^h|\boldsymbol{p}_{\tau-1}, s_{\tau-1}^h, \mathbf{\Theta}^h\} \cdot \Pr\{s_1^h|\mathbf{\Theta}^h\}\, dG_t(\tilde{\boldsymbol{p}}). \quad (5)$$

The marginal probability of the loyalty state, $\Pr\{s_t^h|\mathbf{\Theta}^h\}$, directly follows from (5).

In practice, many researchers have avoided estimators based on the likelihood (4) due to the computational burden associated with estimating $\Pr\{s_t^h|\mathbf{\Theta}^h\}$, especially when covariates are included in the model. The extant literature has typically evaluated the initial state in one of two ways:

1. Assume (i) initial stationarity of the process and (ii) that the stationary distribution of initial loyalty states, $\boldsymbol{\pi}(\mathbf{\Theta}^h)$, does not depend on $\mathbf{\Theta}^h$. Then drop the first observed choice for each household h to initialize
$$s_1^h = \begin{cases} j & \text{if } y_0^h = j \text{ and } j \neq 0, \\ s_0^h & \text{otherwise} \end{cases}$$
[e.g., Seetharaman et al., 1999, Shum, 2004].

2. Set the initial state to zero, $s_1^h = 0$, implying that there is no loyalty to any of the alternatives in the first observed purchase period [e.g., Keane, 1997, Dubé et al., 2010].

Both approaches are expected to generate biased estimates of $\boldsymbol{\theta}$ for samples with finite length T. [3]

For the first approach, the bias arises from the ignored dependence of s_1^h on unobserved consumer tastes. Intuitively, the household's initial state is selected on the preference parameter vector $\mathbf{\Theta}^h$, creating bias if there are idiosyncratic preference components that are unobserved to the

[3] Another approach that has been proposed in the literature is to assume initial stationarity of the process, estimate $\boldsymbol{\pi}(\mathbf{\Theta}^h)$ with auxiliary data, and draw the initial state as part of the MCMC procedure, $s_1^h \sim \boldsymbol{\pi}(\mathbf{\Theta}^h)$. While computationally convenient, such auxiliary data would need to be found on a case-by-case basis [e.g., Sudhir and Yang, 2014], limiting the generalizability of this approach.

researcher. Assuming that s_1^h is exogenous is likely to create an upward bias in the estimate of γ, since the estimator will attribute some of the persistence in choices to s_1^h and not to the preference components, β_j.

The second approach is expected to generate attenuation bias in γ due to the misspecification of the initial condition, which is assumed to be $s_1^h = 0$. The estimator is likely to attribute some of the persistence in choices to the preference components β_j and not to the state dependence.

2.2 Bayesian MCMC Estimation

In this section we describe the MCMC estimator we use to derive the posterior distribution of the taste parameters for the multinomial logit model with state dependence described above in section 2.1. We then describe how we modify the MCMC algorithm to specify the initial conditions. We use the unconditional (on the initial condition) likelihood (4). This approach is similar to Heckman [1981b]'s frequentist estimator that integrates the initial state variable out of the likelihood. Here, we apply a Bayesian approach that treats the initial state as an auxiliary variable.

2.2.1 Exogenous initial conditions

We begin with the case where the initial condition is observed and exogenous. This approach gives rise to the standard, random coefficients logit discrete choice model. In addition to the initial conditions problem discussed in section 2.1, we also address the well-known challenge of separating heterogeneity and state dependence. To ensure that we fully control for heterogeneity in tastes between households, we follow Rossi et al. [2005] and use a hierarchical prior. At the first stage, we specify a K-component mixture of normals distribution for $\left\{\boldsymbol{\Theta}^h\right\}_{h=1}^H$. At the second stage, we specify priors on the parameters of the mixture of normals:

$$\boldsymbol{\Theta}^h | ind_h, \{\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\} \sim N(\boldsymbol{\mu}_{ind_h}, \boldsymbol{\Sigma}_{ind_h}) \tag{6}$$

$$ind_h \sim MN(\boldsymbol{\lambda})$$

$$\boldsymbol{\lambda}, \{\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\} | \boldsymbol{b} \sim \mathbf{F^b},$$

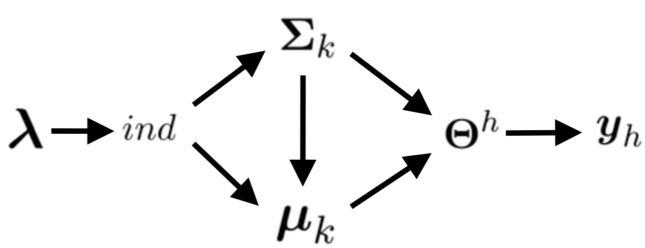
where $\cdot|\cdot$ are conditional distributions and $\boldsymbol{b}$ comprises all the hyperparameters of the priors for both the mixing probabilities and the mixture components. To facilitate the MCMC algorithm, we also introduce the notation $ind_h \in \{1, \dots, K\}$, the latent variable indicating the mixing component from which an individual's preferences are drawn. We specify ind_h as a K-category multinomial random variable with outcome probabilities $\boldsymbol{\lambda} = (\lambda_1, ..., \lambda_K)$. Following Rossi et al. [2005], we specify Dirichlet, normal, and Inverse-Wishart priors on $\boldsymbol{\lambda}$, $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$, respectively, which we denote $\mathbf{F^b}$, and use the same prior hyperparameter settings as in Dubé et al. [2010]. To sample from the resulting posterior distribution, we use a hybrid MCMC procedure with a customized

Metropolis chain [Rossi et al., 2005] for the draws of $\mathbf{\Theta}^h$ and a standard Gibbs sampler for the draws of ind_h, $\boldsymbol{\lambda}$ and $\{\boldsymbol{\mu}_k, \mathbf{\Sigma}_k\}$. The sampling process is

1. $\{\boldsymbol{\mu}_k, \mathbf{\Sigma}_k\}|ind, \mathbf{\Theta}$
2. $ind|\boldsymbol{\lambda}, \{\boldsymbol{\mu}_k, \mathbf{\Sigma}_k\}, \mathbf{\Theta}^h$
3. $\boldsymbol{\lambda}|ind$
4. $\mathbf{\Theta}^h|\{\boldsymbol{\mu}_k, \mathbf{\Sigma}_k\}, ind^h, \{\boldsymbol{y}_h\}$

The first step consists of drawing $\{\boldsymbol{\mu}_k, \mathbf{\Sigma}_k\}$ from a multivariate regression model conditional on the draws of the components, ind, and draws of consumer preferences, $\mathbf{\Theta}$. The second step consists of drawing the components ind^h conditional on $\boldsymbol{\lambda}$, the prior probability of membership in each component, $\mathbf{\Theta}^h$, the consumer-level preferences, and $\{\boldsymbol{\mu}_k, \mathbf{\Sigma}_k\}$, with the multinomial probabilities in the form of likelihood ratios weighted by $\boldsymbol{\lambda}$. The third step consists of drawing $\boldsymbol{\lambda}$ conditional on the updated mixture memberships, ind. Finally, the fourth step consists of the Metropolis step to draw $\mathbf{\Theta}^h$ conditional on the distribution of consumer preferences, $\{\boldsymbol{\mu}_k, \mathbf{\Sigma}_k\}$, the component draw ind^h, and the observed data for consumer h, $\{\boldsymbol{y}_h\}$.[4] Figure 1 provides a DAG for the sampling procedure.

Figure 1: Sampling schema without the initial state draws.



Schema of the Hierarchical Multinomial Logit with a mixture of normals as the first-stage prior.

2.2.2 Extension to include Pr $\{s_1|\boldsymbol{p}_1, ..., \boldsymbol{p}_T, \mathbf{\Theta}\}$

We treat s_1 as an auxiliary variable, adding a draw from s_1 into the MCMC sampling procedure. Throughout this section, we assume that the choice process is initially in equilibrium.

In the absence of prices, it is straightforward to compute the steady state probabilities of each

[4] Our set-up closely follows Rossi et al. [2005], and we refer the interested reader to it for details.

loyalty state. Let

$$\boldsymbol{P}(\boldsymbol{\Theta}) = \begin{bmatrix} \Pr\{s_{t+1}=1|s_t=1,\boldsymbol{\Theta}\} & \cdots & \Pr\{s_{t+1}=J|s_t=1,\boldsymbol{\Theta}\} \\ \vdots & & \vdots \\ \Pr\{s_{t+1}=1|s_t=J,\boldsymbol{\Theta}\} & \cdots & \Pr\{s_{t+1}=J|s_t=J,\boldsymbol{\Theta}\} \end{bmatrix} \tag{7}$$

denote the matrix of transition probabilities. Since the choice process consists of an ergodic, finite-state Markov chain, the distribution over loyalty states converges to the unique stationary distribution $\boldsymbol{\pi}(\boldsymbol{\Theta})$: $\lim_{t\to\infty} \boldsymbol{P}(\boldsymbol{\Theta})^t = \boldsymbol{\iota}\boldsymbol{\pi}(\boldsymbol{\Theta})$, where $\boldsymbol{\iota}$ is a J-vector of ones, $\boldsymbol{\pi}(\boldsymbol{\Theta})$ is a row vector, and the state-specific probabilities, $\boldsymbol{\pi}(\boldsymbol{\Theta}) = (\pi_1(\boldsymbol{\Theta}), ..., \pi_J(\boldsymbol{\Theta}))$, are strictly positive. We can then solve for the steady-state probabilities, $\boldsymbol{\pi}(\boldsymbol{\Theta}) = \boldsymbol{\pi}(\boldsymbol{\Theta})\boldsymbol{P}(\boldsymbol{\Theta})$. [5]

The incorporation of prices into the choice process complicates the computation of the steady state. Since the initial prices in period zero, $\boldsymbol{p}_0$, are typically unobserved, the distribution over the initial state has the form

$$\boldsymbol{\pi}(\boldsymbol{p}_1, ..., \boldsymbol{p}_T, \boldsymbol{\Theta}) = (\pi_1(\boldsymbol{p}_1, ..., \boldsymbol{p}_T, \boldsymbol{\Theta}), ..., \pi_J(\boldsymbol{p}_1, ..., \boldsymbol{p}_T, \boldsymbol{\Theta})). \tag{8}$$

The dependence on the observed prices, $\boldsymbol{p}_1, ..., \boldsymbol{p}_T$, arises because these prices are in general informative about $\boldsymbol{p}_0$. For instance, if the price process is Q-order history-dependent, then $\boldsymbol{p}_0$ depends on the prices in periods $t = 1 \dots Q$. The price $\boldsymbol{p}_0$ affects the choice in period 0 and thus the realized loyalty state in period 1.

In the remainder of the paper, we will assume that prices have support in a discrete set $\{\boldsymbol{\zeta}_1, \dots, \boldsymbol{\zeta}_N\}$.[6]

Below we describe the sampling procedure of $\Pr\left\{s_1|\boldsymbol{p}_1, ..., \boldsymbol{p}_T, \boldsymbol{\Theta}^h\right\}$ for the cases of i.i.d. and Markov price processes. We propose simulation-based solutions to the initial conditions problem that incorporates a draw of s_1 into the MCMC procedure:

$$s_1^h \sim \boldsymbol{\pi}\left(\boldsymbol{p}_1, ..., \boldsymbol{p}_T, \boldsymbol{\Theta}^h\right). \tag{9}$$

Figure 2 provides a DAG for the adjusted sampling procedure.

2.2.3 Price Process is i.i.d.

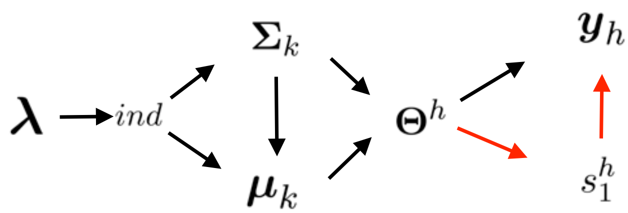
When prices are i.i.d., the unobserved value of $\boldsymbol{p}_0$ does not depend on the observed prices, $\boldsymbol{p}_1, ..., \boldsymbol{p}_T$, simplifying the marginal distribution of the initial states to $\boldsymbol{\pi}(\boldsymbol{\Theta}) = (\pi_1(\boldsymbol{\Theta}), ..., \pi_J(\boldsymbol{\Theta}))$.

To compute $\boldsymbol{\pi}(\boldsymbol{\Theta})$, we use the following approach:

[5] $\boldsymbol{\pi}(\boldsymbol{\Theta})$ is the left eigenvector of $\boldsymbol{P}$ corresponding to the eigenvalue 1.

[6] For similar CPG categories to the ones studied herein, Eichenbaum et al. [2011, page 239] show that "weekly prices typically fluctuate between reference and non-reference prices." In our analysis, we will consider the case where prices are dependent and independent over time.

Figure 2: Sampling schema with the initial state draws.



Schema of the Hierarchical Multinomial Logit with a mixture of normals as the first-stage prior and auxiliary initial state s_0.

1. Use a frequency estimator to estimate the discrete distribution of prices, $\Pr\{\boldsymbol{p}_t = \boldsymbol{\zeta}_n\} = \rho_n,\ n = 1, ..., N$.

2. For each price state $\boldsymbol{\zeta}_n$ and loyalty state s_t, compute $\Pr\left\{s_{t+1}|\boldsymbol{p}_t = \boldsymbol{\zeta}_n, s_t, \boldsymbol{\Theta}^h\right\}$, using equations (1) and (2).

3. Compute the transition probability of s, unconditional on $\boldsymbol{p}_0$:

$$\Pr\left\{s_{t+1} = j|s_t = k, \boldsymbol{\Theta}^h\right\} = \sum_{n=1}^{N} \Pr\left\{s_{t+1} = j|\boldsymbol{p}_t = \boldsymbol{\zeta}_n, s_t = k, \boldsymbol{\Theta}^h\right\}\rho_n.$$

4. Compute the marginal distribution of loyalty states, $\boldsymbol{\pi}(\boldsymbol{\Theta}^h)$, which is a left eigenvector of the transition probability matrix with elements $\Pr\left\{s_{t+1}|s_t, \boldsymbol{\Theta}^h\right\}$.

2.2.4 Price Process is Markov

When prices follow a Markov process, calculating the marginal distribution of the initial state is more complicated because now the state will depend on the observed prices, $\boldsymbol{p}_1, ..., \boldsymbol{p}_T$, as explained above. We assume that prices follow a first-order Markov process, an assumption that has frequently been made in the literature. Note that conditional on $\boldsymbol{p}_1$, the prices $\boldsymbol{p}_2, ..., \boldsymbol{p}_T$ provide no additional information about $\boldsymbol{p}_0$ in the case of a first-order Markov process. Hence, the distribution over the initial loyalty state conditional on the observed sequence of prices, $\boldsymbol{\pi}\left(\boldsymbol{p}_1, \boldsymbol{\Theta}\right)$, depends on $\boldsymbol{p}_1$ only.

We use the following approach to compute the initial distribution of the loyalty state:

1. Estimate the transition probabilities of the price process, $\Pr\left\{\boldsymbol{p}_{t+1} = \boldsymbol{\zeta}_n|\boldsymbol{p}_t = \boldsymbol{\zeta}_j\right\} = \rho_{jn}$, using a frequency estimator. If the price process is ergodic, we compute the stationary distribution $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_N)$,where $\rho_n = \Pr\{\boldsymbol{p}_t = \boldsymbol{\zeta}_n\}$.

2. For each household h, simulate the price chain backwards for periods $t = 0, -1, \ldots, -S$. The draws are obtained assuming that the price process is in equilibrium, which implies the conditional probability distribution:

$$\Pr\{\boldsymbol{p}_{t-1} = \boldsymbol{\zeta}_n | \boldsymbol{p}_t = \boldsymbol{\zeta}_j\} = \frac{\rho_n \rho_{nj}}{\rho_j}.$$

3. Draw a state s_{-S}^h in period $t = -S$ from an exogenous distribution (in the simulations we assume that s_{-S}^h is uniformly distributed over the domain of possible states). Conditional on the price draws obtained in step 2, $\boldsymbol{p}_{-S}, \ldots, \boldsymbol{p}_0$, use the conditional choice probabilities to simulate a sequence of choices and states to obtain a draw s_1^h.

4. Repeat steps 2 and 3 L times, and obtain an estimate of $\boldsymbol{\pi}\left(\boldsymbol{p}_1, \boldsymbol{\Theta}^h\right)$ based on the empirical frequency

$$\pi_j^h = \frac{1}{L} \sum_{l=1}^{L} \mathbb{1}\{s_{1l}^h = j\}.$$

3 Monte Carlo Experiments

We now conduct a series of Monte Carlo experiments to assess the degree of bias due to the initial conditions problem. Of particular interest is the relationship between the bias and the dimensions of a consumer panel database.

3.1 Design

As in section 2.1, we assume that consumer h obtains the following indirect utility from purchasing product j at date t:

$$u_j(\boldsymbol{p}_t, s_t^h, \boldsymbol{\epsilon}_t^h; \boldsymbol{\Theta}^h) = \begin{cases} \beta_j^h - \alpha^h p_{jt} + \gamma^h \mathbb{1}\{s_t^h = j\} + \epsilon_{jt}^h & \text{if } j \neq 0, \\ \epsilon_{0t}^h & \text{if } j = 0. \end{cases}$$

p_{jt} is the price of the product j, and the parameter vector is $\boldsymbol{\Theta}^h = \left(\beta_1^h, ..., \beta_J^h, \alpha^h, \gamma^h\right)'$.

We use the following random coefficients specification to accommodate the persistent heterogeneity in tastes between consumers:

$$\boldsymbol{\Theta}^h \sim N\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right). \tag{10}$$

$\boldsymbol{\sigma} = (\sigma_{\beta_1}, ..., \sigma_{\beta_J}, \sigma_\alpha, \sigma_\gamma)$ denotes the population standard deviations ($\sigma_l = \sqrt{\boldsymbol{\Sigma}_{ll}}$).

To calibrate the parameter values, we use demand estimates based on the Nielsen-Kilts consumer panel purchase data for the margarine category. We discuss the estimates from the Nielsen-Kilts data in section 4.2, and describe the panel data in Appendix 6.1. Accordingly, we set $J = 4$ and use the following hyper-parameter settings for the population distribution of tastes:

$$\begin{bmatrix} \beta_1^h \\ \beta_2^h \\ \beta_3^h \\ \beta_4^h \\ \alpha^h \\ \gamma^h \end{bmatrix} \sim N\left(\begin{bmatrix} -1.71 \\ 0.44 \\ -1.37 \\ -0.91 \\ -1.23 \\ 1 \end{bmatrix}, \begin{bmatrix} 10.37 & 0 & 0 & 0 & 0 & 0 \\ 0 & 10.49 & 0 & 0 & 0 & 0 \\ 0 & 0 & 8.24 & 0 & 0 & 0 \\ 0 & 0 & 0 & 17.22 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1.9 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \right). \tag{11}$$

We assume that the price process is independent of the product choices of individual consumers and estimate it using the empirical distribution of the observed prices in the Nielsen data. Accordingly, we estimate the first-order Markov price process $\Pr\left\{\boldsymbol{p}_{t+1} = \boldsymbol{\zeta}_n | \boldsymbol{p}_t = \boldsymbol{\zeta}_j\right\} = \rho_{jn}$, with prices discretized into $N = 100$ mass points using k-mean clustering of the observed price vectors.With this design, the shares of shopping trips during which brands 1-4 are purchased are 14%, 22%, 16%, and 17%, respectively. The outside option is chosen on 31% of the trips.

$\mathcal{D}$ denotes the data used for estimation. For each synthetic dataset, we simulate a panel for $1,000$ time periods (or trips to the store) and $H \in \{500, 2000\}$ households. To ensure that the simulated data are not influenced by the initialization of the price and choice processes, we drop the first 100 simulation periods and retain the remaining $T = 900$ periods for each household.

We compare four estimation approaches that differ based on the respective treatment of the initial conditions:

1. Retain the first observation for each household and set $s_1^h = 0$.

2. Include the conditional choice probability in the first period assuming that the initial state, s_1^h, is observed, but ignore the marginal probability of the initial loyalty state, $\Pr\left\{s_1^h | \boldsymbol{\Theta}^h\right\}$, in the likelihood function.

3. Assume that the price process is i.i.d., and draw $s_1^h \sim \boldsymbol{\pi}(\boldsymbol{\Theta}^h)$ as an auxiliary variable that is drawn from the stationary distribution of loyalty states, as described in section 2.2.3.

4. Treat the price process as first-order Markov, and draw the auxiliary variable s_1^h from $\boldsymbol{\pi}\left(\boldsymbol{p}_1, \boldsymbol{\Theta}^h\right)$, as described in section 2.2.4.

For each of the four estimation approaches we perform $R = 100$ independent replications, and we estimate each of the model specifications for each replication.We report summary statistics of the posterior mean of the first-stage prior, i.e. random coefficients distribution, of the state dependence parameter, γ. First, for replication r and draw d we calculate

$$\boldsymbol{\mu}_{dr} = \sum_{k=1}^{K} \lambda_{k,dr} \boldsymbol{\mu}_{k,dr},$$
$$\boldsymbol{\Sigma}_{dr} = \sum_{k=1}^{K} \lambda_{k,dr} \boldsymbol{\Sigma}_{k,dr} + \sum_{k=1}^{K} \lambda_{k,dr} (\boldsymbol{\mu}_{k,dr} - \boldsymbol{\mu}_{dr})(\boldsymbol{\mu}_{k,dr} - \boldsymbol{\mu}_{dr})'.$$

Here, $\lambda_{k,dr}$ is the d^{th} draw of the mixture probability for component k, and $\boldsymbol{\mu}_{k,dr}$ and $\boldsymbol{\Sigma}_{k,dr}$ are the corresponding draws of the mean and variance of component k.[7] We then average over all D MCMC draws to obtain the posterior mean and variance of the first-stage prior in replication r:

$$\hat{\boldsymbol{\mu}}_r = \frac{1}{D} \sum_{d=1}^{D} \boldsymbol{\mu}_{dr},$$
$$\hat{\boldsymbol{\Sigma}}_r = \frac{1}{D} \sum_{d=1}^{D} \boldsymbol{\Sigma}_{dr}.$$

$\hat{\mu}_{\gamma,r}$ is the γ-component of $\hat{\boldsymbol{\mu}}_r$, and $\hat{\sigma}_{\gamma,r}$ is the square root of the γ-component of the diagonal of $\hat{\boldsymbol{\Sigma}}_r$.

We report the mean and standard deviation of $\hat{\mu}_{\gamma,r}$ and $\hat{\sigma}_{\gamma,r}$ across the R replications, and we also report the corresponding 2.5^{th} and 97.5^{th} percentiles. Finally, we report the percentage of replications where the true values of μ_γ and σ_γ are contained in the respective 95% posterior credible intervals.

3.2 Monte Carlo Results

3.2.1 Main Results

In this section we contrast the various approaches to specifying the initial conditions and their respective abilities to recover the true preference parameters in the presence of state dependence ($\gamma^h > 0$). We discuss the small-sample properties of γ^h for panels that vary in the length of T.

Our first experiment uses a design intended to mimic a panel length of approximately 6 years, which is much longer than the panels used in the extant literature.[8] For instance, in the 6-year period between 2006 and 2011 of the Nielsen-Kilts database discussed below in section 4.2, we observe a mean, median and maximum number of observed trips per household of $T = 22$, $T = 14$ and $T = 220$ respectively.[9] Hence, for our first experiment, we use a balanced panel with $T = 30$ for

[7] We show the formulas for $\boldsymbol{\mu}_{k,dr}$ and $\boldsymbol{\Sigma}_{k,dr}$ for the case of arbitrarily many mixture components. If $K = 1$, the formulas simplify in an obvious manner.

[8] For example, Keane [1997] used all trips over a 60-week period, Dubé et al. [2010] use a subset of all trips over a 120-week period, and Seetharaman et al. [1999] use all the weeks in a 3-year period.

[9] Recall that T is based in part on our definition of the outside good, which we define as a purchase of any

Table 1: Monte Carlo Results: Posterior mean and standard deviation estimates of γ. $T = 30$ and $N = 2000$.

		Mean of $\hat{\mu}_{\gamma,r}$ and $\hat{\sigma}_{\gamma,r}$ (in brackets: coverage) (2.5, 97.5 quantiles across the MC draws)			
Point Estimates	True Value	(1) $s_0 = 0$	(2) $p(s_0\|\theta)$ ignored	(3) $p(s_0\|\theta)$ included Prices i.i.d.	(4) Prices Markov
Mean $\hat{\mu}_{\gamma,r}$	1.0	0.759 (0)	3.044 (0)	1.102 (0.68)	1.066 (0.86)
		(0.678, 0.819)	(2.673, 3.345)	(0.997, 1.202)	(0.967, 1.163)
Mean $\hat{\sigma}_{\gamma,r}$	1.0	0.852 (0)	2.817 (0)	1.007 (0.93)	0.977 (0.91)
		(0.77, 0.899)	(2.496, 3.094)	(0.905, 1.092)	(0.881, 1.052)

Based on a 100 Monte Carlo simulations.

each household, which is a long panel compared to the CPG datasets used in the extant literature.

In Table 1, we compare the results for each of the 4 model specifications. Columns (1) and (2) present the estimates of the mean $\hat{\mu}_{\gamma,r}$ and $\hat{\sigma}_{\gamma,r}$ for the cases when the initial condition is set to $s_1^h = 0$ and when $\Pr\left\{s_1^h|\mathbf{\Theta}^h\right\}$ is not included in the likelihood, respectively. The estimates of both $\hat{\bar{\gamma}}$ and $\hat{\sigma}_\gamma$ are biased. Across our 100 replications, the true values of μ_γ and σ_γ never lie inside the 95% credible interval. Moreover, μ_γ and σ_γ lie outside the $(2.5^{th}, 97.5^{th})$ percentile range of the point estimates across the simulations.

Columns (3) and (4) present the estimates of mean $\hat{\mu}_{\gamma,r}$ and $\hat{\sigma}_{\gamma,r}$ for the cases where we correct the likelihood of the initial conditions to account for both the choice and price processes; although column (3) makes the incorrect simplifying assumption that the price process is i.i.d.. Across the 100 replications, the estimates are close to the true values of μ_γ and σ_γ, which fall within the $(2.5^{th}, 97.5^{th})$ percentile range of the point estimates and over half the 95% credible intervals. Interestingly, the point estimates for the i.i.d. price scenario (column 3) and Markov price scenario (column 4) are qualitatively similar, even though the true price process is simulated as Markov. However, using the correct Markov specification on prices slightly improves the coverage of $\hat{\mu}_{\gamma,r}$.[10]

In sum, we find that even with a simulated panel that would be deemed "long enough" based on past research, we still find bias in the two specifications that exhibit specification error in the likelihood of the initial conditions. To understand the source of this bias, first recall that consumers select the outside option on 31% of the trips. So when $T = 30$, the average consumer makes only 22 purchases and the number of observed transitions in the loyalty state has a mean and median of only 3.87 and 2 respectively. Moreover, the loyalty state never deviates from its initial value for 45.4% of the consumers in our simulated sample. Consequently, even with $T = 30$, we may not

other product in the margarine category. This reduces the number of outside option choices in the data and makes the simulations more efficient. All results presented below qualitatively hold if we define the outside option as any purchase of a spreadable product or as any trip to a store, after adjusting T accordingly.

[10] Since coverage is computed using only 100 Monte Carlo simulations, differences in the coverage of $\hat{\sigma}_{\gamma,r}$ in specifications (3) and (4) can occur by chance.

observe enough state transitions to avoid a small-sample bias due to initial conditions.[11] In Table 9 of Appendix 6.3, we further show that even for an unrealistically long panel with all consumers having $T = 100$ observations, specifications (1) and (2) still lead to a significant (although smaller in magnitude) bias, while specifications (3) and (4) correctly recover $\hat{\mu}_{\gamma,r}$ and $\hat{\sigma}_{\gamma,r}$.

3.2.2 Results By Number of Purchases

To explore when a dataset is capable of eliminating the initial conditions bias, we next run experiments that balance the panel based on the number of observed purchases, T^{purch}, instead of on the number of purchase occasions (i.e., trips). We draw the data set for $T = 1,000$ time periods for N consumers, remove the first 100 simulated choices and then retain T^{purch} trips per consumer, truncating after T^{purch} purchases are observed.[12] Following standard practices in the extant brand choice literature, we exclude consumers who do not make at least T^{purch} choices in T periods. As a result, our simulation sample is slightly less than N across the simulation conditions, which we report with the results. In Table 3, we compare the results for several scenarios that re-balance the panel on $T^{purch} \in \{5, 10, 25, 100\}$ to examine the role of the number of observed purchases. We also vary $N \in \{500, 2000\}$ to examine the role of the size of the cross-section of consumers used.

This common approach to constructing a panel with a sufficient number of purchases selects on consumers' unobserved tastes. It is therefore even more crucial in this scenario to verify whether, even in very long panels, it is possible to recover the true parameter values in practice.

Since the loyalty parameter is identified by the observed spells in the data, we summarize the number of spells for each of our scenarios in Table 2. For realistic values of $T^{purch} \leq 10$, which is approximately consistent with typical purchase panel lengths used in practice, we observe very few spells per consumer (e.g., less than 3 on average) and over half the panelists fail to deviate from their initial loyalty state. It is only when we increase T^{purch} well above these realistic scenarios that we see considerably more spells. When $T^{purch} = 100$, almost 80% of the panelists deviate from their initial state at least once and the average panelist has 16.2 spells. When $T^{purch} = 100$, the mean and median panel length across simulated consumers is $T = 176.3$ and $T = 112$, respectively, corresponding to well over a decade of purchases per household.

Table 3 presents the Monte Carlo results. Column (1) presents the estimates of the mean of $\hat{\mu}_{\gamma,r}$ and $\hat{\sigma}_{\gamma,r}$ when the initial state is set to zero (i.e. ignore initial conditions in first observed period). Confirming the results in Table 1, mean $\hat{\mu}_{\gamma,r}$ and $\hat{\sigma}_{\gamma,r}$ are systematically *underestimated*, across all the T^{purch} and N conditions. For the case when $N = 2,000$ households, the true in-sample values of μ_γ and σ_γ are almost never covered by the 95% credible interval, even when each panelist

[11] Changing the definition of the outside good to increase the number of observations (e.g., to include more of the observed store trips) would not remedy this problem since it would only increase the number of non-purchase observations without changing the number of observed state transitions.

[12] We remove the first 100 simulated choices for each consumer to make sure that the initial state in the data is drawn from the stationary distribution of consumer choices.

Table 2: Switching Behavior in a Typical Simulated Panel

T^{purch}	Spells		% panelists that
	mean	median	deviate from s_0
5	1.68	1	34.5%
10	2.47	1	48.2%
25	4.97	3	62%
100	16.2	9	78%

makes $T^{purch} = 100$ purchases, corresponding to on average 176.3 trips. For the case of $N = 500$ households, the coverage is slightly higher due to a higher variance in the posterior point estimates, but still well below our criterion of 95%. The true in-sample values of μ_γ and σ_γ lie outside of the $(2.5^{th}, 97.5^{th})$ percentile range across the point estimates in each replication for most of the cases. Although, when $T^{purch} = 100$, μ_γ is just inside the range and σ_γ is just outside the range, suggesting that increasing T^{purch} does indeed reduce the bias.

Column (2) of table 3 presents the estimates when the initial state is set correctly but treated as exogenous. In this case, the mean of $\hat{\mu}_{\gamma,r}$ and $\hat{\sigma}_{\gamma,r}$ are *overestimated* for all but the largest values of T^{purch}. For small values of $T^{purch} \in \{5, 10\}$, the true in-sample values of μ_γ and σ_γ are never covered by the 95% credible interval, and the $(2.5^{th}, 97.5^{th})$ percentile range of the point estimates across replications is shifted substantially to the right of the true values. Once again, the bias appears to decline as we increase T^{purch}. For $T^{purch} = 25$, the bias is small and for $T^{purch} = 100$, the bias appears to have dissipated.

Columns (3) and (4) of table 3 present the estimates of the mean of $\hat{\mu}_{\gamma,r}$ and $\hat{\sigma}_{\gamma,r}$ with the initial state drawn as an auxiliary variable in the MCMC routine. For all T^{purch} and N specifications, the recovered estimates are close to the true in-sample values of μ_γ and σ_γ. The $(2.5^{th}, 97.5^{th})$ percentile range of the point estimates across replications also contain the true values. However, it is only when we use larger values of T^{purch} that we start to see the true values falling within the 95% posterior credible intervals for at least 90% of the replications. Interestingly, while our price process was simulated as Markov, our treatment of prices in the MCMC sampling either as i.i.d. or Markov has little effect on the resulting point estimates. Using the correct Markov specification on prices only slightly improves the coverage when $N = 500$, for instance.

In sum, specification (1) leads to a substantial attenuation bias in γ, specification (2) overestimates the mean and variance of γ, and specifications (3) and (4) correctly recover the state dependence distribution in the majority of the cases. When $T^{purch} = 5$, the relatively low performance in terms of percentage of replications where the posterior credible intervals contain the true values is due to the fact that many households do not switch brands with so few observed purchases.

In practice, researchers do not truncate their purchase panels above at a maximum number of purchases. Rather, a researcher would likely pre-screen their sample to ensure that retained

Table 3: Monte Carlo Results: Posterior mean and standard deviation estimates of γ. Balanced panel.

			Mean of $\hat{\mu}_{\gamma,r}$ and $\hat{\sigma}_{\gamma,r}$ (in brackets: coverage) (2.5, 97.5 quantiles across the MC draws)			
N	T^{Purch}	Point Estimates (Means)	(1) $s_0 = 0$	(2) $p(s_0\|\theta)$ ignored	(3) $p(s_0\|\theta)$ included Prices i.i.d.	(4) Prices Markov
$N = 2,000$:						
	5	$\hat{\mu}_\gamma$	**0.43** (0)	**4.17** (0)	1.30 (0.32)	**1.04** (0.87)
			(0.31, 0.55)	(3.92, 4.43)	(1.08, 1.58)	(0.83, 1.29)
		$\hat{\sigma}_\gamma$	**0.62** (0)	**2.81** (0)	**1.33** (0.15)	**1.05** (0.93)
			(0.56, 0.69)	(2.52, 3.08)	(1.15, 1.56)	(0.86, 1.26)
	10	$\hat{\mu}_\gamma$	**0.62** (0)	**3.7**9 (0)	**1.06** (**0.85**)	**0.98** (0.89)
			(0.54, 0.69)	(3.54, 4.02)	(0.94, 1.21)	(0.88, 1.1)
		$\hat{\sigma}_\gamma$	**0.7** (0)	**3.02** (0)	**1.05** (0.88)	**0.98** (0.94)
			(0.64, 0.77)	(2.83, 3.22)	(0.93, 1.16)	(0.88, 1.06)
	25	$\hat{\mu}_\gamma$	**0.79** (0)	**1.45** (0.09)	**1.02** (0.93)	**1** (0.89)
			(0.73, 0.86)	(1.07, 2.54)	(0.93, 1.1)	(0.93, 1.07)
		$\hat{\sigma}_\gamma$	**0.85** (0)	**1.41** (0.05)	**1** (0.9)	**0.99** (0.92)
			(0.79, 0.91)	(1.04, 2.48)	(0.93, 1.07)	(0.92, 1.05)
	100	$\hat{\mu}_\gamma$	**0.96** (0)	**1.06** (0.94)	**1.06** (0.96)	**1.05** (0.97)
			(0.92, 1)	(1.01, 1.13)	(1.01, 1.12)	(1, 1.11)
		$\hat{\sigma}_\gamma$	**0.94** (0.06)	**1** (0.92)	**1** (0.95)	**0.99** (0.91)
			(0.89, 0.97)	(0.95, 1.05)	(0.95, 1.04)	(0.94, 1.03)
$N = 500$:						
	5	$\hat{\mu}_\gamma$	**0.56** (0.05)	**4.26** (0)	**1.54** (0.33)	**1.41** (0.56)
			(0.34, 0.84)	(3.73, 4.82)	(1.09, 1.98)	(0.95, 1.91)
		$\hat{\sigma}_\gamma$	**0.81** (0.6)	**2.96** (0)	**1.44** (0.33)	**1.27** (0.75)
			(0.7, 0.94)	(2.57, 3.47)	(1.15, 1.87)	(0.99, 1.64)
	10	$\hat{\mu}_\gamma$	**0.67** (0.01)	**3.71** (0)	**1.15** (0.81)	**1.07** (0.95)
			(0.52, 0.83)	(3.35, 4.13)	(0.88, 1.41)	(0.84, 1.27)
		$\hat{\sigma}_\gamma$	**0.78** (0.15)	**2.92** (0)	**1.1** (0.89)	**1.03** (0.96)
			(0.67, 0.9)	(2.62, 3.23)	(0.91, 1.33)	(0.86, 1.2)
	25	$\hat{\mu}_\gamma$	**0.81** (0.03)	**1.78** (0.23)	**1.05** (0.91)	**1.02** (0.94)
			(0.68, 0.93)	(1.02, 3.27)	(0.88, 1.22)	(0.86, 1.18)
		$\hat{\sigma}_\gamma$	**0.86** (0.24)	**1.70** (0.23)	**1.01** (0.91)	**0.99** (0.94)
			(0.76, 0.96)	(0.98, 3)	(0.87, 1.16)	(0.87, 1.12)
	100	$\hat{\mu}_\gamma$	**0.96** (0.37)	**1.08** (0.88)	**1.07** (0.92)	**1.06** (0.93)
			(0.85, 1.06)	(0.95, 1.18)	(0.95, 1.17)	(0.95, 1.16)
		$\hat{\sigma}_\gamma$	**0.95** (0.65)	**1.02** (0.89)	**1.01** (0.93)	**1** (0.95)
			(0.88, 1.02)	(0.94, 1.11)	(0.93, 1.1)	(0.93, 1.09)

Point estimates are the means of $\hat{\sigma}_{\gamma,r}$ and $\hat{\sigma}_{\gamma,r}$ across r.

panelists always have at least some minimum number of purchases. For instance, Honore and Kyriazidou [2000] show that the identification of the state-dependent choice model using a fixed effects approach to capture persistent, unobserved heterogeneity requires $T^{purch} \geq 3$. To replicate this approach, we repeat the same Monte Carlo design as above but simulate the number of purchases per household from the empirical distribution of purchases based on Nielsen-Kilts data, truncated at T^{purch}.[13] In this scenario, T^{purch} is a lower bound on the number of observed purchases per panelist.

When we use T^{purch} as a lower bound, we mechanically obtain more information than in the balanced scenario. At the low end, $T^{purch} \leq 5$ generates a mean and median number of spells of 3.61 and 2.0 respectively, although 49% of panelists still fail to deviate from their initial loyalty state. At the same time, we now have 446 households with at least 20 purchases. Of interest is whether pooling households with short purchase histories contaminates our final estimates even though we have a substantial number of households with long histories. At the high end of $T^{purch} \geq 100$, the mean and median number of spells is 19.3 and 10 respectively, and only 22% of panelists fail to deviate from the initial loyalty state. However, as before, $T^{purch} \geq 100$ corresponds to an unrealistically long panel length, implying a mean and median number of observed trips of $T = 195$ and $T = 144$ respectively.

We report the results from this second set of experiments in Table 4. Most of our findings are qualitatively similar to those in Table 3. The key difference is that our corrected procedures (columns 3 and 4) now perform even better, likely because we have more brand-switching in these data.

There are several important take-aways from these results. First and perhaps most important, we find systematic short-sample biases in γ, the coefficient on the loyalty state, when we use a model that mis-specifies the likelihood over the initial states, even with a relatively wide cross section ($N = 2,000$) and long panel ($T^{purch} = 100$). Interestingly, in the unbalanced case, where we observe many purchases for most consumers, the inclusion of consumers with few (e.g., $T^{purch} < 10$) purchases (a typical case for CPG categories) still biases the estimates.[14] In this sense, the results in Table 4 when $T^{purch} \geq 5$ and $N = 500$ are close to what we expect in an application of demand estimation with state dependence in a CPG category. Columns (3) and (4) of Table 4 show that we obtain largely unbiased estimates once we account for the likelihood of the initial conditions.

Second, the directions of the bias in specifications (1) and (2) allow us to bound the mean of the state dependence coefficient, μ_γ, without estimating the more computationally complex procedures that account for the likelihood of the initial conditions. When we set the initial states to 0 (specification 1), the estimates of μ_γ are downward biased. Intuitively, setting $s_0 = 0$ forces the estimation procedure to infer that the first purchase of each consumers is fully driven by consumers'

[13] The histogram of the empirical distribution of purchases from Nielsen-Kilts data is presented in Appendix 6.2,

[14] In the Nielsen HMS data that we use later, the share of consumers with 3-5 purchases is 37.2%, with the median number of purchases of 7 and mean of 12.7.

Table 4: Monte Carlo Results: Posterior mean and standard deviation estimates of γ. Unbalanced panel.

			Mean of $\hat{\mu}_{\gamma,r}$ and $\hat{\sigma}_{\gamma,r}$ (in brackets: coverage) (2.5, 97.5 quantiles across the MC draws)			
N	T^{Purch}	Point Estimates (Means)	(1) $s_0 = 0$	(2) $p(s_0\|\theta)$ ignored	(3) $p(s_0\|\theta)$ included Prices i.i.d.	(4) Prices Markov
$N = 2000$:						
	5	$\hat{\mu}_\gamma$	**0.71** (0)	**3.58** (0)	**1.03** (0.92)	**1.01** (0.95)
			(0.63, 0.78)	(3.35, 3.9)	(0.94, 1.12)	(0.91, 1.1)
		$\hat{\sigma}_\gamma$	**0.81** (0)	**3.14** (0)	**1.01** (0.91)	**0.99** (0.96)
			(0.75, 0.89)	(2.93, 3.34)	(0.92, 1.12)	(0.92, 1.09)
	10	$\hat{\mu}_\gamma$	**0.79** (0.05)	**2.73** (0)	**1.04** (0.92)	**1.03** (0.96)
			(0.73, 0.87)	(2.25, 3.06)	(0.96, 1.13)	(0.95, 1.1)
		$\hat{\sigma}_\gamma$	**0.85** (0.02)	**2.61** (0)	**1.01** (0.94)	**0.99** (0.98)
			(0.78, 0.9)	(2.15, 2.88)	(0.95, 1.09)	(0.93, 1.07)
	25	$\hat{\mu}_\gamma$	**0.87** (0)	**1.09** (0.42)	**1.04** (0.94)	**1.02** (0.95)
			(0.82, 0.92)	(1.02, 1.18)	(0.98, 1.1)	(0.96, 1.08)
		$\hat{\sigma}_\gamma$	**0.89** (0.01)	**1.05** (0.53)	**1** (0.95)	**0.98** (0.92)
			(0.84, 0.93)	(0.99, 1.11)	(0.95, 1.05)	(0.93, 1.03)
	100	$\hat{\mu}_\gamma$	**0.97** (0)	**1.07** (0.92)	**1.06** (0.95)	**1.05** (0.94)
			(0.92, 1.02)	(1.01, 1.13)	(1.01, 1.12)	(1.1, 1.11)
		$\hat{\sigma}_\gamma$	**0.94** (0.13)	**1** (0.94)	**1** (0.97)	**0.99** (0.96)
			(0.9, 0.98)	(0.95, 1.05)	(0.95, 1.04)	(0.95, 1.03)
$N = 500$:						
	5	$\hat{\mu}_\gamma$	**0.75** (0.05)	**3.52** (0)	**1.08** (0.83)	**1.04** (0.9)
			(0.6, 0.91)	(3.03, 3.94)	(0.86, 1.3)	(0.85, 1.25)
		$\hat{\sigma}_\gamma$	**0.85** (0.36)	**3.05** (0)	**1.03** (0.93)	**1.01** (0.91)
			(0.71, 1)	(2.62, 3.37)	(0.84, 1.24)	(0.83, 1.2)
	10	$\hat{\mu}_\gamma$	**0.8** (0.04)	**2.56** (0.05)	**1.05** (0.86)	**1.03** (0.92)
			(0.67, 0.92)	(1.16, 3.39)	(0.87, 1.25)	(0.85, 1.19)
		$\hat{\sigma}_\gamma$	**0.86** (0.32)	**2.41** (0.06)	**1.01** (0.92)	**1** (0.93)
			(0.75, 0.95)	(1.1, 3.07)	(0.86, 1.15)	(0.84, 1.13)
	25	$\hat{\mu}_\gamma$	**0.87** (0.12)	**1.12** (0.64)	**1.05** (0.91)	**1.03** (0.92)
			(0.74, 0.99)	(0.97, 1.31)	(0.91, 1.21)	(0.9, 1.18)
		$\hat{\sigma}_\gamma$	**0.91** (0.42)	**1.08** (0.68)	**1.01** (0.92)	**1** (0.92)
			(0.78, 1.01)	(0.9, 1.28)	(0.87, 1.15)	(0.87, 1.14)
	100	$\hat{\mu}_\gamma$	**0.97** (0.41)	**1.07** (0.92)	**1.06** (0.95)	**1.06** (0.98)
			(0.88, 1.1)	(0.96, 1.22)	(0.95, 1.2)	(0.95 1.2)
		$\hat{\sigma}_\gamma$	**0.94** (0.66)	**1** (0.95)	**1** (0.94)	**0.99** (0.95)
			(0.87, 1.01)	(0.92, 1.09)	(0.92, 1.09)	(0.91, 1.07)

T^{purch} is now a lower bound on the number of observed purchases per retained panelist. Point estimates are the means of $\hat{\sigma}_{\gamma,r}$ and $\hat{\sigma}_{\gamma,r}$ across r.

persistent preferences for the product, β, and price sensitivity, α. In practice, however, the first observed choice is also determined by the true initial state. Correspondingly, there is attenuation bias in the μ_γ estimate. Conversely, when we drop the initial choice and correctly define the initial state but treat it as exogenous (specification 2), μ_γ is overestimated. Now the estimation routine assumes that the consumer has arrived in the initial state by pure chance, ignoring the fact that the initial state is a function of the consumer's preferences. As a result, any subsequent choices of the consumer that match the initial state are attributed to the state dependence process and not to consumer's preferences, leading to an upward bias in the μ_γ estimate.

3.3 A Diagnostic Tool

The Monte Carlo results in section 3 suggest a simple method for diagnosing whether or not a state dependence correction is necessary when analyzing a given choice-panel dataset. While we do not derive an analytic characterization of the biases associated with the various approaches for handling the specification of the initial conditions, the results herein are suggestive of a computationally simple approach for bounding the degree of state dependence in a purchase panel. To bound the true value of μ_γ, one can use the two ad hoc estimators:

1. **Lower bound**: retain the first observation for each household and set $s_1^h = 0$
2. **Upper bound**: drop the first observation so that s_1^h is observed, but ignore $\Pr\left\{s_1^h|\mathbf{\Theta^h}\right\}$ in the likelihood.

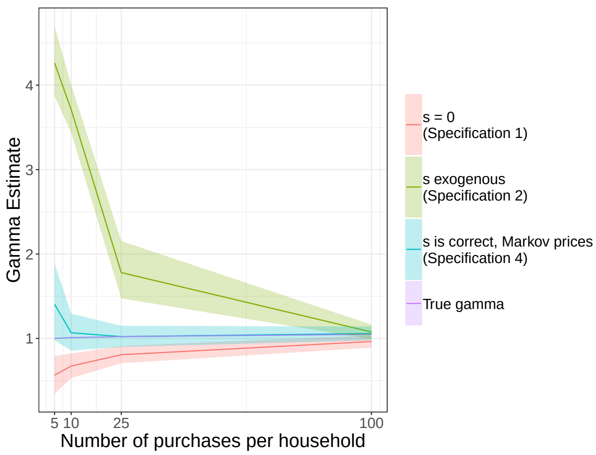
If the bounds are tight, there will be no benefit to estimating the more computationally intensive, corrected approaches that we proposed.

In Figure 3, we plot the mean of $\hat{\mu}_{\gamma,r}$ across Monte Carlo replications along with the corresponding (average) credible intervals. We report these statistics for both the balanced and unbalanced panels, for panel size $N \in \{500, 2000\}$ and length $T^{purch} \in \{5, 10, 25, 100\}$. The bounds are wide for low values of T^{purch} and narrow as T^{purch} increases. The correct value always lies within the bounds. At $T^{purch} = 100$, the difference between the bounds is small, with an expected difference of 0.103 (9.8% of the true μ_γ).

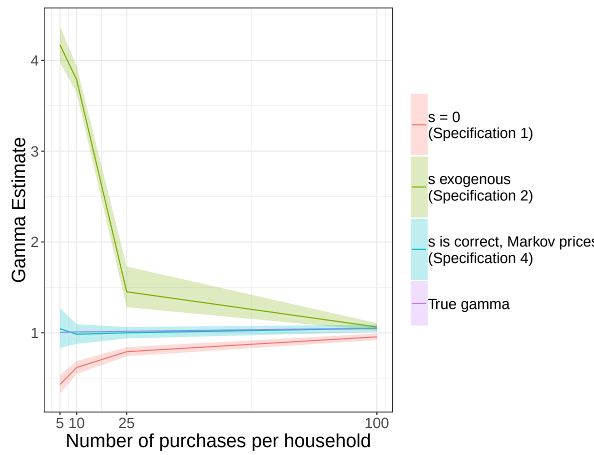
In Figure 4 we conduct a similar exercise using a new Monte Carlo design with no state dependence: $\mu_\gamma = \sigma_\gamma = 0$. Subfigure (a) presents the average values of $\hat{\mu}_{\gamma,r}$ in our main specifications for different levels of T^{purch}. As expected, almost all specifications recover the true value of μ_γ.[15] Subfigure (b) plots the range of the bounds implied by the estimates of the specifications (1) and (2) with its corresponding 95% posterior credible interval. For all T^{purch}, we cannot reject that the range of the bounds is zero (in other words, the difference in the mean of $\hat{\mu}_{\gamma,r}$ between specifications

[15] A notable exception is the upper bound specification with $T^{purch} = 5$. In this case, state dependence is slightly overestimated (the estimate is around 0.069 with a 95% posterior credible interval of (0.02, 0.119). This result shows that the model struggles to separate out state dependence and heterogeneity when T^{purch} is small.

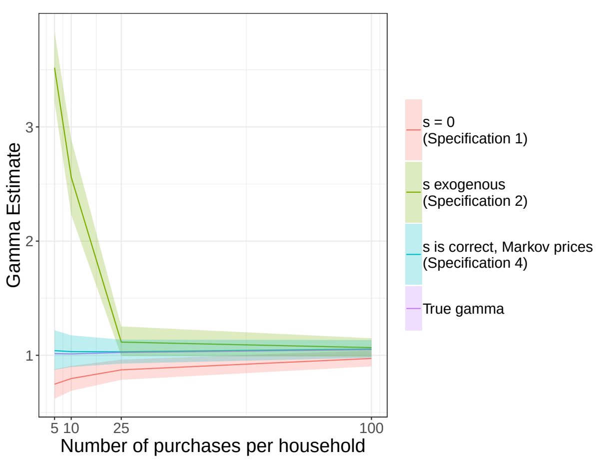
Figure 3: Mean $\hat{\mu}_{\gamma,r}$ and the corresponding average credible intervals as a function of T.



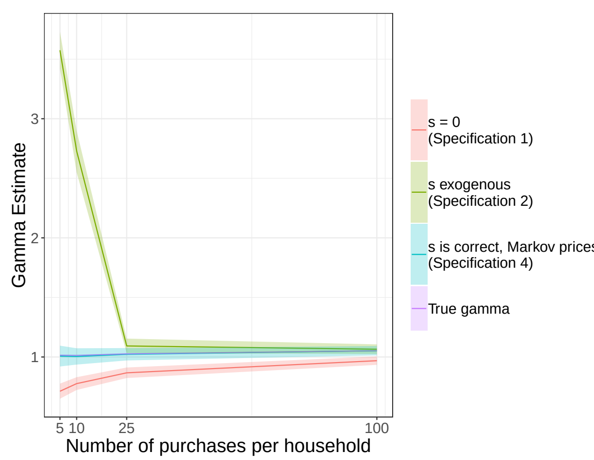
(a) Balanced, N = 500



(b) Balanced, N = 2,000



(c) Unbalanced, N = 500

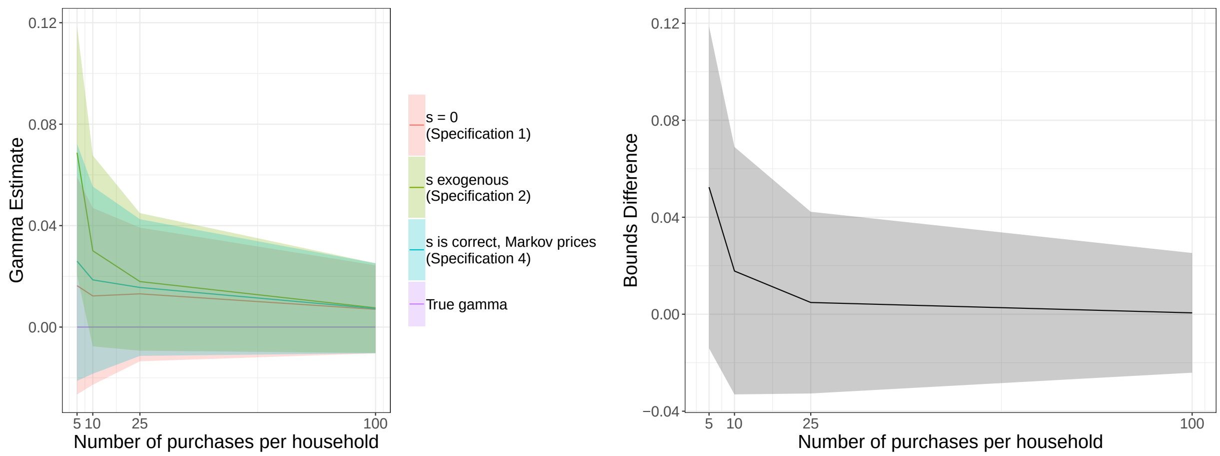


(d) Unbalanced, N = 2,000

Shaded region represents an average of 2.5% and 97.5% quantiles across the Monte Carlo replications. "True gamma" almost exactly coincides with specification 4 in all subfigures.

1 and 2 is not statistically significant), implying that the initial conditions correction method is not necessary.[16]

Figure 4: Mean $\hat{\mu}_{\gamma,r}$ and the implied bounds for the simulation case of $\gamma = 0$.



(a) Average values of $\hat{\mu}_{\gamma,r}$

(b) Bounds on μ_γ

Subfigure (a): Mean $\hat{\mu}_{\gamma,r}$ and the corresponding average credible intervals as a function of T; Subfigure (b): Bounds on μ_γ implied by specifications (1) and (2). Shaded regions represent an average of 2.5% and 97.5% quantiles across the Monte Carlo replications. Case of the unbalanced panel and N = 2,000.

Overall, our findings indicate that the computationally simple ad hoc approaches can be used as a preliminary check for positive state dependence. In the absence of state dependence, these estimators appear to generate the correct bound of zero even for short panels. In the presence of state dependence, the bounds converge as the panel length increases, potentially providing a diagnostic check for whether a given dataset is sufficiently long to be able to ignore the initial conditions without biasing the parameter estimates.

4 Empirical Application

The simulations in Section 3 demonstrates how the misspecification of the likelihood of the initial conditions can bias demand estimates, even in moderately long panels. To illustrate this problem in practice, we now estimate the same four specifications from Section 3 using two actual consumer purchase panels for the margarine category. The panels are from the Denver Scantrack Data in Dubé et al. [2010] and from the Nielsen-Kilts Homescan Data for 2006-2011.

[16] Appendix 6.4 explores the robustness of our findings to data-generating processes with different magnitudes of state-dependence: $\mu_\gamma \in \{-1, 0.5\}$. The bounds apply to $\mu_\gamma = 0.5$ but not to the process with negative state dependence on average, $\mu_\gamma = -1$. Appendix 6.4 discusses the intuition behind this result.

4.1 Denver Scantrack Data

Our first implementation uses the data in Dubé et al. [2010]. The data contain 429 households making tub margarine purchases in the Denver Scantrack between January 1993 and March 1995. The choice set comprises the four leading margarine brands. On average, households make 16.7 purchases from this choice set (ranging from 3 to 92) over the sample period. We refer the interested reader to Dubé et al. [2010] for a detailed description of the data and the estimation sample.

We estimate the price process as follows. We discretize the support of the vector of observed prices charged by the four brands by classifying the observed prices into 100 clusters, using the cluster centers as prices states. We then estimate the transition probability matrix of prices across the states using the frequency estimator. To account for the frequency of shopping trips of consumers, we compute the median number of days between the shopping trips of consumers in the Nielsen data (4 days) and simulate the price process assuming that consumers observe prices (make shopping trips) every 4 days. Although not reported herein, all our estimated state transitions have positive probabilities and therefore our empirical price process satisfies the required ergodicity condition.[17]

Table 5 reports the posterior expectation of the mean and standard deviation of the first-stage prior, $\hat{\mu}$ and $\hat{\sigma}$, for each of the four specifications discussed in section 3.1. As expected, we observe large differences in our parameter estimates for specifications (1) and (2), which treat the initial conditions as exogenous, versus specifications (3) and (4), which model the dependence of the initial states on unobserved consumer tastes. As in the Monte Carlo experiments, setting the initial states to zero (specification 1) leads to much smaller estimates of $\hat{\mu}_\gamma$ and $\hat{\sigma}_\gamma$, than the specifications that incorporate the likelihood of the initial states (specifications 3 and 4). Treating the initial states as exogenous (specification 2) leads to much higher estimates of $\hat{\mu}_\gamma$ and $\hat{\sigma}_\gamma$ than the specifications that model the likelihood of the initial states (specifications 3 and 4). These differences are significant in the sense that a 95% posterior credible interval of the difference in the posterior expected point estimate does not contain zero. Interestingly, the estimates in specifications (3) and (4) are very close, and the similarity in their posterior likelihoods are too close to allow us to reject one model in favor of the other.

We also observe corresponding differences in some of the other preference parameters. However, there does not appear to be any systematic pattern in these differences.

These findings are consistent with our Monte Carlo experiments. The differences between our first two specifications and second two specifications are broadly consistent with the biases documented in section 3. These findings suggest that the estimated levels of loyalty documented in Dubé et al. [2010] underestimate the actual degree of loyalty in the margarine category.[18]

[17] The estimates for the price process are available upon request.

[18] This bias does not change the finding that equilibrium prices are lower due to switching costs in Dubé et al. [2009], who find that prices only start to rise when switching costs are four times the magnitude of the empirical estimates.

Table 5: Estimation results with Denver Scantrack Data. Estimates of μ and σ across the estimation specifications.

		Posterior point estimates of $\mu$and σ (95 % posterior credible interval)			
	$\hat{\mu}_l$	(1)	(2)	(3)	(4)
	and	$s_0 = 0$	$p(s_0\|\theta)$ ignored	$p(s_0\|\theta)$ included	
Price:	$\hat{\sigma}_l$			i.i.d.	Markov
Brand 1	$\hat{\mu}_{\beta_1}$	1.967	1.417	1.165	1.091
		(1.129, 2.862)	(0.36, 2.363)	(0.111, 2.131)	(0.157, 2.004)
	$\hat{\sigma}_{\beta_1}$	2.039	1.177	1.667	1.917
		(1.26, 3.153)	(0.701, 1.889)	(1.013, 2.951)	(1.102, 3.297)
Brand 2	$\hat{\mu}_{\beta_2}$	-3.077	-0.33	-2.629	-2.928
		(-4.767, -1.447)	(-1.446, 0.667)	(-4.408, -1.169)	(-4.464, -1.738)
	$\hat{\sigma}_{\beta_2}$	5.016	1.425	3.354	3.797
		(3.37, 7.038)	(0.86, 2.124)	(1.935, 5.001)	(2.587, 5.882)
Brand 3	$\hat{\mu}_{\beta_3}$	0.247	-0.294	-0.297	-0.358
		(-0.382, 0.872)	(-1.012, 0.372)	(-1.08, 0.356)	(-1.047, 0.318)
	$\hat{\sigma}_{\beta_3}$	3.66	1.51	3.125	3.335
		(2.836, 4.64)	(1.06, 2.021)	(2.258, 4.224)	(2.357, 4.509)
Brand 4	$\hat{\mu}_{\beta_4}$	2.329	1.662	1.433	1.349
		(1.575, 3.186)	(0.658, 2.544)	(0.438, 2.346)	(0.403, 2.248)
	$\hat{\sigma}_{\beta_4}$	2.288	1.466	2.029	2.155
		(1.508, 3.345)	(0.95, 2.147)	(1.325, 2.962)	(1.252, 3.433)
Price	$\hat{\mu}_{\alpha}$	-2.656	-2.744	-2.269	-2.227
		(-3.204, -2.162)	(-3.287, -2.134)	(-2.838, -1.661)	(-2.749, -1.653)
	$\hat{\sigma}_{\alpha}$	1.52	0.982	1.213	1.333
		(0.972, 2.353)	(0.71, 1.392)	(0.729, 1.941)	(0.811, 2.231)
State Dependence	$\hat{\mu}_{\gamma}$	**0.667**	**3.096**	**1.125**	**1.138**
		(0.494, 0.836)	(2.847, 3.379)	(0.852, 1.428)	(0.868, 1.44)
	$\hat{\sigma}_{\gamma}$	**0.984**	**2.73**	**1.324**	**1.319**
		(0.797, 1.182)	(2.436, 3.041)	(1.06, 1.635)	(1.034, 1.625)

Hierarchical Multinomial Logit with Normal Heterogeneity. 100,000 MCMC draws.

4.2 Nielsen-Kilts Data

We now repeat the analysis from Section 4.1 using the Nielsen-Kilts Homescan data. The advantage of these data is that we can construct a much longer choice panel to see if the biases documented above persist when a longer choice sequence is available to the researcher.

We construct our purchase panel by combining the Nielsen-Kilts Homescan and RMS datasets.[19] For details on the construction of the estimation sample see Appendix 6.1. The sample contains 1,829 households that constitute at least 85% of their observed margarine purchases in a single chain observed in the RMS database. To mimic the design of the Denver sample used above, we focus our analysis on 15 oz and 16oz tubs of margarine, the top-selling pack size, and the four top-selling brands: "Imperial," "I Can't Believe Its not Butter," "Blue Bonnet," and "Smart Balance." We collapse the resulting 31 UPCs into four brand "items." In the final estimation sample, we observe a mean and median number of purchases per household of 12.66 and 7 respectively.

As before, we estimate each of the four specifications. Table 6 reports the posterior expectation of $\hat{\mu}$ and $\hat{\sigma}$ for each of the four specifications. We replicate all the key substantive conclusions from the Denver Scantrack data. Thus, the results from specifications (3) and (4), which address the dependence of the initial conditions on unobserved consumer tastes in the likelihood, generate similar estimates. In Table 7, we report the differences in posterior estimates of the mean and standard deviation of γ between our first two specifications, which likely have misspecified initial conditions, and specification (4) which simulates the likelihood of the initial conditions. Consistent with our Monte Carlo evidence, we find that specification (1) yields a significantly smaller estimate (in a statistical sense) than specification (4), whereas specification (2) is significantly larger. In both comparisons, we reject the hypothesis of equal μ_γ and σ_γ values with at least 95% posterior probability.

An advantage of the Nielsen-Kilts data is the fact we can track some households for a relatively long period. To investigate the role of "big T^{purch}," we estimate the four specifications using only those households for which we observe at least 10 purchases, leaving us with a sample of 713 households. We report the estimates in Table 8. As in our Monte Carlo experiments, we still observe differences between specifications (1) and (2) relative to (3) and (4). However, as we increase T^{purch}, while the differences in the estimates of the distribution of γ in specifications 1 and 2 become smaller, there is still a detectable bias. In addition, the estimates based on the truncated sample are substantially different from the full sample estimates (Table 6).

5 Conclusions

While the initial conditions problem associated with the state-dependent choice processes has been discussed in the econometrics literature since at least Heckman [1981b], most of the empirical

[19] We use RMS data to get the product prices in a given store during the week of the household's purchase.

Table 6: Estimation results with Nielsen HMS Data. Estimates of μ and σ across the estimation specifications.

		Posterior point estimates of $\mu$and σ (95 % posterior credible interval)			
	$\hat{\mu}_l$ and $\hat{\sigma}_l$	(1) $s_0 = 0$	(2) $p(s_0\|\theta)$ ignored	(3) $p(s_0\|\theta)$ included	(4)
Price:				i.i.d.	Markov
Brand 1	$\hat{\mu}_{\beta_1}$	-1.703	-2.157	-2.056	-1.984
		(-1.896, -1.528)	(-2.361, -1.958)	(-2.253, -1.861)	(-2.161, -1.815)
	$\hat{\sigma}_{\beta_1}$	3.235	2.966	3.391	3.292
		(3.02, 3.452)	(2.77, 3.176)	(3.153, 3.62)	(3.063, 3.505)
Brand 2	$\hat{\mu}_{\beta_2}$	0.442	-0.152	0.033	0.04
		(0.185, 0.686)	(-0.456, 0.133)	(-0.221, 0.279)	(-0.201, 0.291)
	$\hat{\sigma}_{\beta_2}$	3.222	2.992	3.304	3.306
		(2.924, 3.529)	(2.696, 3.31)	(3.014, 3.59)	(3.002, 3.612)
Brand 3	$\hat{\mu}_{\beta_3}$	-1.384	-1.977	-1.618	-1.679
		(-1.539, -1.238)	(-2.169, -1.791)	(-1.772, -1.475)	(-1.832, -1.521)
	$\hat{\sigma}_{\beta_3}$	2.911	2.705	2.818	2.938
		(2.718, 3.107)	(2.516, 2.912)	(2.624, 3.008)	(2.739, 3.135)
Brand 4	$\hat{\mu}_{\beta_4}$	-0.907	-0.828	-1.031	-1.104
		(-1.24, -0.572)	(-1.165, -0.505)	(-1.364, -0.718)	(-1.441, -0.783)
	$\hat{\sigma}_{\beta_4}$	4.139	3.173	3.875	3.992
		(3.801, 4.499)	(2.863, 3.504)	(3.511, 4.239)	(3.656, 4.332)
Price	$\hat{\mu}_{\alpha}$	-1.238	-1.204	-1.143	-1.149
		(-1.343, -1.133)	(-1.323, -1.089)	(-1.241, -1.045)	(-1.253, -1.04)
	$\hat{\sigma}_{\alpha}$	1.375	1.234	1.375	1.374
		(1.255, 1.507)	(1.117, 1.362)	(1.249, 1.494)	(1.233, 1.498)
State Dependence	$\hat{\mu}_{\gamma}$	**0.651**	**2.635**	**1.06**	**1.026**
		(0.582, 0.72)	(2.491, 2.777)	(0.952, 1.178)	(0.932, 1.122)
	$\hat{\sigma}_{\gamma}$	**0.909**	**2.608**	**1.194**	**1.148**
		(0.84, 0.98)	(2.445, 2.788)	(1.078, 1.313)	(1.057, 1.254)

(Hierarchical Multinomial Logit with Normal Heterogeneity. 100,000 MCMC draws.)

Table 7: Posterior estimates of the differences in γ across specifications (95% posterior credible intervals in brackets).

$\hat{\mu}_{\gamma}^{(1)} - \hat{\mu}_{\gamma}^{(4)}$	$\hat{\mu}_{\gamma}^{(2)} - \hat{\mu}_{\gamma}^{(4)}$	$\hat{\sigma}_{\gamma}^{(1)} - \hat{\sigma}_{\gamma}^{(4)}$	$\hat{\sigma}_{\gamma}^{(2)} - \hat{\sigma}_{\gamma}^{(4)}$
-0.375	1.609	-0.239	1.46
(-0.483,-0.265)	(1.434,1.789)	(-0.359,-0.121)	(1.263,1.664)

Table 8: Estimation results with Nielsen HMS Data using only panelists with 10+ purchases. Estimates of μ and σ across the estimation specifications.

		Posterior point estimates of $\mu$and σ (95 % posterior credible interval)			
	$\hat{\mu}_l$	(1)	(2)	(3)	(4)
	and	$s_0 = 0$	$p(s_0\|\theta)$ ignored	$p(s_0\|\theta)$ included	
Price:	$\hat{\sigma}_l$			i.i.d.	Markov
Brand 1	$\hat{\mu}_{\beta_1}$	-1.335	-1.479	-1.631	-1.552
		(-1.608, -1.088)	(-1.839, -1.158)	(-1.848, -1.381)	(-1.811, -1.338)
	$\hat{\sigma}_{\beta_1}$	3.324	3.272	3.484	3.418
		(3.006, 3.584)	(2.912, 3.607)	(3.043, 3.758)	(2.972, 3.736)
Brand 2	$\hat{\mu}_{\beta_2}$	0.464	0.395	0.079	0.156
		(0.176, 0.747)	(0.053, 0.714)	(-0.221, 0.397)	(-0.169, 0.514)
	$\hat{\sigma}_{\beta_2}$	3.239	3.144	3.384	3.399
		(2.851, 3.566)	(2.758, 3.586)	(2.98, 3.766)	(3.052, 3.734)
Brand 3	$\hat{\mu}_{\beta_3}$	-1.258	-1.345	-1.381	-1.422
		(-1.453, -1.059)	(-1.565, -1.105)	(-1.577, -1.192)	(-1.649, -1.215)
	$\hat{\sigma}_{\beta_3}$	3.164	2.934	2.961	3.121
		(2.868, 3.438)	(2.451, 3.298)	(2.618, 3.224)	(2.762, 3.411)
Brand 4	$\hat{\mu}_{\beta_4}$	-0.889	-0.684	-0.994	-1.089
		(-1.289, -0.354)	(-1.039, -0.303)	(-1.375, -0.621)	(-1.51, -0.691)
	$\hat{\sigma}_{\beta_4}$	4.084	3.472	3.853	4.008
		(3.415, 4.529)	(3.075, 3.883)	(3.403, 4.293)	(3.622, 4.377)
Price	$\hat{\mu}_{\alpha}$	-1.095	-1.116	-1.01	-1.027
		(-1.227, -0.967)	(-1.246, -0.981)	(-1.129, -0.868)	(-1.174, -0.872)
	$\hat{\sigma}_{\alpha}$	1.426	1.36	1.424	1.439
		(1.257, 1.582)	(1.181, 1.541)	(1.254, 1.579)	(1.228, 1.59)
State Dependence	$\hat{\mu}_{\gamma}$	0.792	1.414	1.124	1.061
		(0.69, 0.996)	(1.044, 2.115)	(0.993, 1.276)	(0.957, 1.252)
	$\hat{\sigma}_{\gamma}$	1.017	1.601	1.333	1.241
		(0.931, 1.111)	(1.22, 2.312)	(1.219, 1.437)	(1.132, 1.357)

(Hierarchical Multinomial Logit with Normal Heterogeneity. 100,000 MCMC draws.)

literature on brand choice assumes the panel length is sufficiently large to eliminate the bias created by the improper treatment of the initial condition. In a series of Monte Carlo experiments, we find that initial conditions bias can persist even in panel datasets that are much longer than those typically used in the extant literature. Moreover, popular and simple approaches to initialize the choice process can introduce systematic bias into the estimates of state-dependence. In particular, the common assumption that the first observed choice does not exhibit state dependence leads to substantial under-estimation of the degree of state-dependence in the choice process, while dropping the first observation and treating the initial state as exogenous leads to substantial over-estimation.

Under the assumption of stationary initial conditions, we employ simulation methods to augment standard Bayesian procedures for choice models with state dependence to correct the initial conditions problem. The method provides much more accurate estimates of the true degree of state dependence, and provides good sampling performance even in panels with only a handful of purchases per panelist.

Our results suggest (though without a formal proof) a simple diagnostic approach to assess whether a state dependence correction is necessary. Researchers can bound the extent of the initial condition bias by running two estimation specifications: one with all initial states set to zero, and one with the observed initial state treated as an exogenous realization. If these two specifications lead to statistically similar estimates, a correction procedure likely would not change the results substantially. However, if the estimates are substantially different, researchers can apply a correction procedure like the one proposed herein. This correction procedure is more computationally intensive but is generally available since it does not require any auxiliary data for the initial choice.

References

G. H. Brown. Brand loyalty fact or fiction. *Advertising Age*, 9:53–55, 1952. 1

Pradeep K Chintagunta. Inertia and variety seeking in a model of brand-purchase timing. *Marketing Science*, 17(3):253–270, 1998. 2.1

Jean-Pierre Dubé, Günter Hitsch, Peter E. Rossi, and Maria Ana Vitorino. Category pricing with state-dependent utility. *Marketing Science*, 27(3):417–429, 2008. 1

Jean-Pierre Dubé, Günter J. Hitsch, and Peter E. Rossi. Do switching costs make markets less competitive? *Journal of Marketing Research*, 46:435–445, 2009. 1, 18

Jean-Pierre Dubé, Günter J. Hitsch, and Peter E. Rossi. State dependence and alternative explanations for consumer inertia. *RAND Journal of Economics*, 41(3):417–445, 2010. 1, 2.1, 2, 2.2.1, 8, 4, 4.1, 4.1

Martin Eichenbaum, Nir Jaimovich, and Sergio Rebelo. Reference prices, costs, and nominal rigidities. *American Economic Review*, 101(1):234–262, 2011. 6

Tülin Erdem, Susumu Imai, and Michael P. Keane. Brand and quantity choice dynamics under price uncertainty. *Quantitative Marketing and Economics*, 1:5–64, 2003. 1

Ronald E. Frank. Brand choice as a probability process. *The Journal of Business*, 35(1):43–56, 1962. 1

James J. Heckman. the incidental parameters problem and the problem of initial conditions in estimating a discrete time-discrete data stochastic process and some monte carlo evidence". In C. Manski and D. McFadden, editors, *Structural Analysis of Discrete Data with Econometric Applications*, chapter 4, pages 179–195. MIT Press, 1981a. 1

James J Heckman. The incidental parameters problem and the problem of initial conditions in estimating a discrete time-discrete data stochastic process. 1981b. 2.1, 2.2, 5

James J. Heckman. Identifying the hand of past: Distinguishing state dependence from heterogeneity. *American Economic Review*, 81(2):75–79, 1991. 1

Igal Hendel and Aviv Nevo. Measuring the implications of sales and consumer inventory behavior. *Econometrica*, 74(6):1637–1673, 2006. 1

Bo E. Honore and Ekaterini Kyriazidou. Panel data discrete choice models with lagged dependent variables. *Econometrica*, 68(4):839–874, 2000. 3.2.2

Michael P. Keane. Modeling heterogeneity and state dependence in consumer choice behavior. *Journal of Business & Economic Statistics*, 15(3):310–327, 1997. 1, 2.1, 2, 8

Peter Rossi, Greg Allenby, and Rob McCulloch. *Bayesian Statistics and Marketing*. John Wiley & Sons, 2005. 2.2.1, 2.2.1, 4

P.B. Seetharaman, Andrew Ainslie, and Pradeep Chintagunta. Investigating household state dependence effects across categories. *Marketing Science*, 36:488–500, 1999. 1, 2.1, 1, 8

Sangwoo Shin, Sanjog Misra, and Dan Horsky. Disentangling preferences and learning in brand choice models. *Marketing Science*, 31(1):115–137, 2012. 1

Matthew Shum. Does advertising overcome brand loyalty? evidence from the breakfast-cereals market. *Journal of Economics and Management Strategy*, 13(2):241–272, 2004. 1, 1

K Sudhir and Nathan Yang. Exploiting the choice-consumption mismatch: A new approach to disentangle state dependence and heterogeneity. 2014. 3

6 Appendices

6.1 Appendix A: Nielsen-Kilts HMS and RMS datasets

We use the Nielsen-Kilts Homescan (HMS) and Retail Measurement System (RMS) datasets to construct a long shopping panel. The Nielsen-Kilts HMS data track CPG purchases for a rolling panel of households between 2004 and 2016. The Nielsen-Kilts RMS data tracks weekly prices and unit volumes sold at the UPC level for about 35,000 stores between 2006 and 2015. We focus on the time span between 2006 and 2011, since most households churn out of the sample within 3 or 4 years.

We analyze shopping behavior in the margarine category, focusing on the 16-oz tubs sold by the top 4 national brands, based on expenditures in the category. The resulting set of brands includes "Imperial," "Blue Bonnet," "Smart Balance" and "I Can't Believe Its Not Butter" (ICBNB), spanning 30 unique UPCs. An analysis of price correlations indicates high correlation between products of the same brand and pack size: the median within-group correlations range from 87.2% and 98.8%. We observe considerably lower correlation between products of the same brands and different packaging types: the median within-brand correlations for Blue Bonnet and Smart Balance are 53.4% and 58.2%, respectively. Therefore, we aggregate the UPCs into 6 "products" based on brand and packaging type (stick or tub): one combination for Imperial and ICBNB and two combinations for Blue Bonnet and Smart Balance.

We restrict our sample to households that make at least 85% of their margarine purchases in the same retail chain. This restriction ensures that the households face the same price process over time, a useful simplification for our proposed approach that estimates the Markov price process. In addition, for each brand, we retain the packaging type with the highest sales.[20] We then retained those stores for which all 4 products are observed. Finally, we retained those households making at least 3 purchases during the sample period.

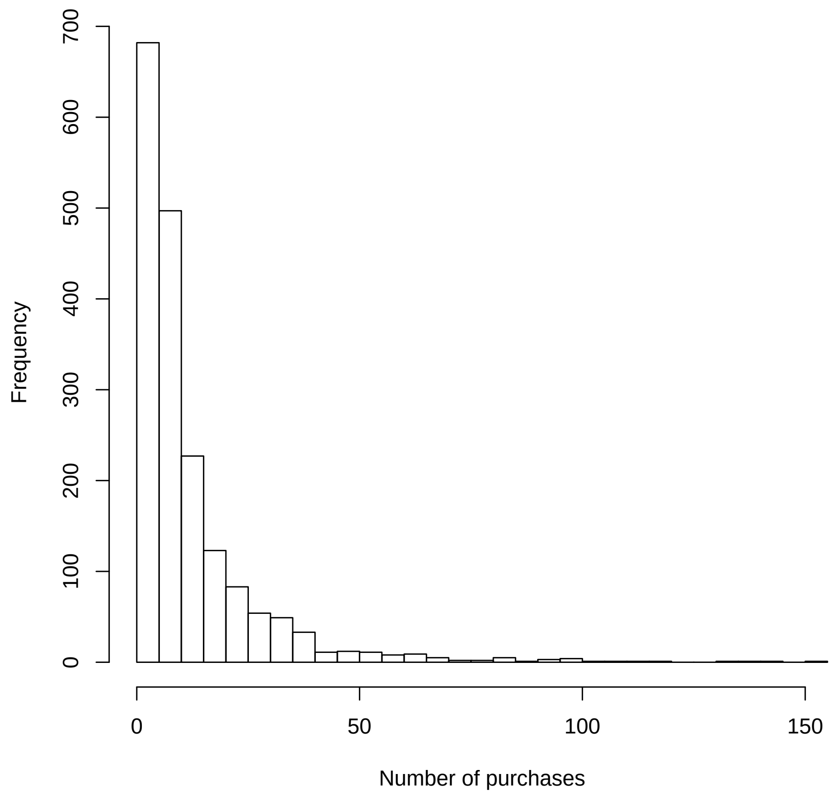
Our final estimation sample comprises 1,829 households making 23,173 purchases from our set of 4 products. In the sample, the mean and median number of purchases per household are 12.7 and 7 respectively. We define the outside option as a purchase of any other margarine product in the store. In contrast with other typical definitions of the outside good (e.g., every trip to the store), this definition reduces the total number of trips in the estimation panel along with the rate at which the outside option is chosen. Households purchase the outside option on 16,998 occasions, or 42.3% of the trips.[21]

[20] The less popular packaging types for Blue Bonnet and Smart Balance each sold one tenth as much as the more popular size and represented only about 2.2% of the purchases in the sample

[21] An alternative way to define an outside option is as a purchase of any spread. In this case the outside option would be chosen on 75.3% of the trips.

6.2 Appendix B: An empirical Distribution of Purchases in Nielsen-Kilts data.

Figure 5: Empirical Distribution of Purchases of Margarine in Nielsen-Kilts data.



6.3 Appendix C: Results with $T = 100$

Table 9: Monte Carlo Results: Posterior mean and standard deviation estimates of γ. $T = 100$ and $N = 2,000$.

		Mean of $\hat{\mu}_{\gamma,r}$ and $\hat{\sigma}_{\gamma,r}$ (in brackets: coverage) (2.5, 97.5 quantiles across the MC draws)			
	Point Estimates	(1) $s_0 = 0$	(2) $p(s_0\|\theta)$ ignored	(3) $p(s_0\|\theta)$ included	(4)
Price:				i.i.d.	Markov
	Mean $\hat{\mu}_{\gamma,r}$	0.869 (0)	**1.075** (0.38)	**1.028** (0.93)	**1.019** (0.99)
		(0.814, 0.918)	(1.006, 1.14)	(0.962, 1.09)	(0.956, 1.077)
	Mean $\hat{\sigma}_{\gamma,r}$	**0.93** (0.06)	**1.023** (0.81)	**0.994** (0.91)	**0.987** (0.92)
		(0.876, 0.979)	(0.958, 1.08)	(0.934, 1.054)	(0.927, 1.043)

Based on a 100 Monte Carlo simulations.

6.4 Appendix D: Robustness of Results to Alternative Values of μ_γ

To check the robustness of our findings in section 3.2 to alternative magnitudes of state dependence, we re-run the Monte Carlo simulations using the values $\mu_\gamma \in \{-1, 0.5\}$, holding all the other true parameter values the same as before. We expect similar qualitative findings for $\mu_\gamma = 0.5$ because this specification still generates positive state dependence. However, the bias results are less clear for $\mu_\gamma = -1$ because when $\gamma < 0$, consumers exhibit variety-seeking. Variety-seeking should cause more consumer brand-switching, which in turn generates more information about state dependence even in a short sample. Moreover, this brand-switching offsets the dependence of the choice history on the initial state. Consequently, we still expect specification (1), which sets the initial state to zero, to generate attenuation bias in the γ estimate. However, the magnitude and direction of bias in specification (2), which treats the first observed loyalty state as exogenous, is unclear.

Figure 6 visualizes the estimates for $\mu_\gamma \in \{-1, 0.5\}$. Subfigure (a) presents the mean of $\hat{\mu}_{\gamma,r}$ estimates when $\mu_\gamma = 0.5$. As expected, specification (1) underestimates the true state dependence and specification (2) overestimates the true state dependence, as we found earlier in section 3.2. More important, our proposed estimator , specification (4), recovers the true value of μ_γ.

Our results differ from those in section 3.2 when $\mu_\gamma = -1$. Specification (1) still under-states the magnitude of state dependence. The bias disappears only once we increase T^{purch} to 100, a very long panel length. Surprisingly, specification (2) does not appear to generate any bias, even when T^{purch} is small. As explained above, this finding may be due to the fact that the increased brand-switching induced by $\gamma < 0$ reduces the dependence of the purchase history on the initial state. Finally, as before, our specification that models the role of the initial conditions recovers the true value of μ_γ. While our interest herein is mainly on loyalty (i.e., positive state dependence), a deeper understanding of initial conditions bias under variety-seeking may be an interesting direction for future research.

Table 10 presents the estimated values of mean of $\hat{\mu}_{\gamma,r}$ and $\hat{\sigma}_{\gamma,r}$. The results provide more information about the estimates presented in Figure 6 – specifications (3) and (4) correctly recover the true distribution of γ, while specification (1) generates biases in both the mean and dispersion. Specification (2) generates biased estimates when $\mu_\gamma = 0.5$ but appears to recover the true distribution of γ when $\mu_\gamma = -1$.[22]

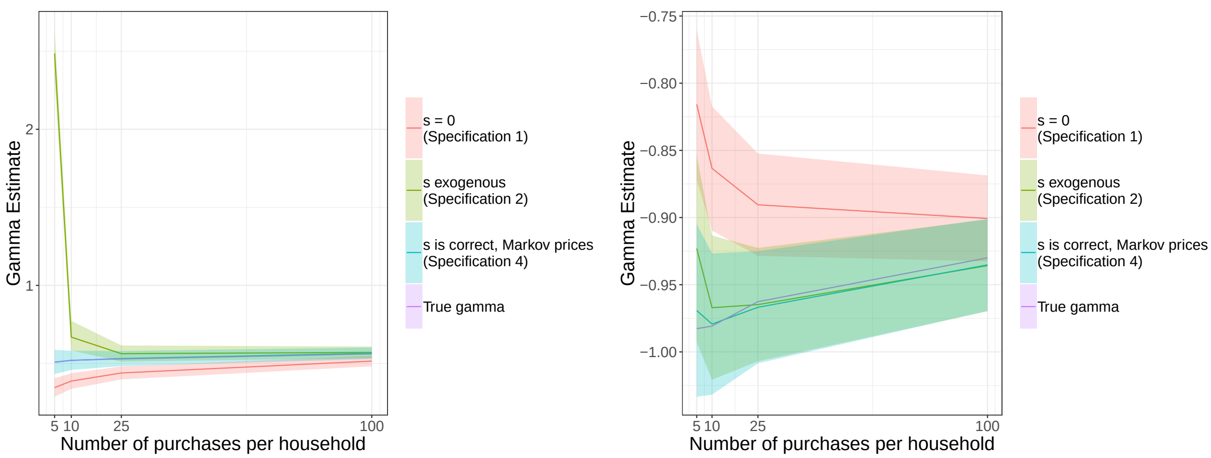
[22] A small exception is the case of $T^{purch} = 5$ when $\mu_\gamma = -1$: the coverage for specification (2) are lower than 95%.

Table 10: Monte Carlo Results: Posterior mean and standard deviation estimates of γ for $\mu_\gamma = 0.5$ and $\mu_\gamma = -1$ specifications. Unbalanced panel, N = 2,000 and $\sigma_\gamma = 1$.

			Mean of $\hat{\mu}_{\gamma,r}$ and $\hat{\sigma}_{\gamma,r}$ (in brackets: coverage) (2.5, 97.5 quantiles across the MC draws)			
μ_γ	T^{Purch}	Point Estimates (Means)	(1) $s_0 = 0$	(2) $p(s_0\|\theta)$ ignored	(3) $p(s_0\|\theta)$ included Prices i.i.d.	(4) Prices Markov
$\mu_\gamma = 0.5$:						
			0.35 (0)	**2.49** (0)	**0.53** (0.88)	**0.51** (0.95)
	5	$\hat{\mu}_\gamma$	(0.29, 0.430)	(2.17, 2.84)	(0.44, 0.63)	(0.44, 0.6)
			0.83 (0)	**2.89** (0)	**1.02** (0.92)	**0.99** (0.94)
		$\hat{\sigma}_\gamma$	(0.76, 0.895)	(2.63, 3.13)	(0.93, 1.1)	(0.91, 1.07)
			0.39 (0)	**0.67** (0.13)	**0.53** (0.95)	**0.52** (0.93)
	10	$\hat{\mu}_\gamma$	(0.33, 0.440)	(0.56, 0.81)	(0.45, 0.61)	(0.45, 0.6)
			0.86 (0)	**1.13** (0.2)	**1** (0.98)	**0.99** (0.95)
		$\hat{\sigma}_\gamma$	(0.81, 0.913)	(1.04, 1.26)	(0.93, 1.07)	(0.93, 1.05)
			0.44 (0.02)	**0.56** (0.69)	**0.54** (0.94)	**0.53** (0.95)
	25	$\hat{\mu}_\gamma$	(0.38, 0.504)	(0.49, 0.64)	(0.47, 0.6)	(0.47, 0.6)
			0.91 (0.03)	**1.03** (0.77)	**1** (0.91)	**1** (0.93)
		$\hat{\sigma}_\gamma$	(0.86, 0.962)	(0.97, 1.09)	(0.95, 1.06)	(0.95, 1.05)
			0.52 (0.25)	**0.57** (0.94)	**0.57** (0.97)	**0.57** (0.95)
	100	$\hat{\mu}_\gamma$	(0.47, 0.563)	(0.52, 0.62)	(0.51, 0.66)	(0.51, 0.61)
			0.95 (0.15)	**1** (0.99)	**0.99** (1)	**0.99** (1)
		$\hat{\sigma}_\gamma$	(0.91, 0.984)	(0.96, 1.04)	(0.95, 1.03)	(0.95, 1.03)
$\mu_\gamma = -1$:						
			-0.82 (0)	**-0.92** (0.63)	**-0.97** (0.92)	**-0.97** (0.91)
	5	$\hat{\mu}_\gamma$	(-0.88, -0.75)	(-1, -0.83)	(-1.04, -0.9)	(-1.04, -0.9)
			0.95 (0.59)	**1.06** (0.69)	**1** (0.91)	**1** (0.9)
		$\hat{\sigma}_\gamma$	(0.87, 1.03)	(0.97, 1.16)	(0.9, 1.08)	(0.92, 1.09)
			-0.86 (0)	**-0.97** (0.92)	**-0.98** (0.90)	**-0.98** (0.91)
	10	$\hat{\mu}_\gamma$	(-0.93, -0.81)	(-1.04, -0.91)	(-1.05, -0.92)	(-1.05, -0.92)
			0.94 (0.43)	**1.02** (0.89)	**1** (0.91)	**1** (0.89)
		$\hat{\sigma}_\gamma$	(0.88, 1)	(0.95, 1.09)	(0.93, 1.07)	(0.93, 1.07)
			-0.89 (0.06)	**-0.97** (0.94)	**-0.97** (0.95)	**-0.97** (0.95)
	25	$\hat{\mu}_\gamma$	(-0.95, -0.83)	(-1.03, -0.9)	(-1.03, -0.91)	(-1.03, -0.9)
			0.96 (0.39)	**1.01** (0.94)	1 (0.9)	**1** (0.89)
		$\hat{\sigma}_\gamma$	(0.9, 1.01)	(0.94, 1.06)	(0.94, 1.06)	(0.94, 1.06)
			-0.9 (0.56)	**-0.94** (0.92)	**-0.94** (0.92)	**-0.94** (0.92)
	100	$\hat{\mu}_\gamma$	(-0.96, -0.85)	(-0.99, -0.88)	(-0.99, -0.88)	(-0.99, -0.88)
			0.98 (0.71)	**1.01** (0.94)	**1.01** (0.91)	**1.01** (0.92)
		$\hat{\sigma}_\gamma$	(0.94, 1.03)	(0.96, 1.05)	(0.95, 1.05)	(0.96, 1.05)

Point estimates are the means of $\hat{\sigma}_{\gamma,r}$ and $\hat{\sigma}_{\gamma,r}$ across r.

Figure 6: Mean of $\hat{\mu}_{\gamma,r}$ for the simulation case of $\mu_\gamma = \{0.5, -1\}$.



(a) $\mu_\gamma = 0.5$ (b) $\mu_\gamma = -1$

Figures present the mean of $\hat{\mu}_{\gamma,r}$ and the corresponding average credible intervals as a function of T. Case of the unbalanced panel and N = 2,000.