Structural State Dependence and Consumers’ Unobserved Heterogeneity: A Case Study of the U.S. Coffee Market

By Ziyi Qiu *
University of Chicago
July 28, 2018

In many markets of differentiated products, consumers show persistence in their consumption choices. Understanding how past consumption affects consumers’ current choices and distinguishing different explanations are beneficial for price setting. In this paper, a discrete choice model is used to investigate the state dependence effect on the U.S. coffee market. Using a disaggregate data set that captures a detailed purchase history and demographic information for a massive number of consumers, I first estimate the spurious state dependence effect without assuming consumer heterogeneity. I then consider both structural state dependence and consumers’ unobserved heterogeneity and distinguish those two effects by applying the discrete mass point algorithm semiparametrically. I further explore the potential correlated errors and endogeneity problem and compare the various methods to solve them. My findings show that structural state dependence and consumers’ unobserved heterogeneity are the main explanations for the observed consumption persistence in the U.S. coffee market. My findings also suggest that there exist two types of consumers in the U.S. coffee market: one type treats ground coffee as normal goods and the other treats ground coffee as inferior goods. The structural state dependence is positive and significant for one type of consumers and negative for the other type of consumers.

I. Introduction

Understanding how consumers make their consumption choices is of broad interest and remains an active research topic in economics and marketing. Researchers seek to understand how consumers select products based on their preferences to maximize utility. Firms benefit from understanding consumers’ behavior in order to set prices, update product attributes, and plan promotion strategies. Most importantly, understanding consumer choices helps economists comprehend the demand system [Ber94, BLP95, Sma87, Nev01, AR02, ABBP06], which has widespread applications for predicting market shares, inferring costs and performing antitrust analysis.

Despite the importance of understanding consumer choices, the pioneer works

* Email: ziyiqiu.uchicago@gmail.com
are limited to study aggregate consumption behavior due to lack of individual-level data. The recent availability of high-quality disaggregate data (from Nielsen and IRI) has pushed the research frontier toward exploring individual consumption choices and identifying individual heterogeneity as a feasible topic. Using disaggregate data, researchers in marketing and economics have documented a form of persistence in consumption choices and have begun to investigate in this problem [Fra62, Mas66, Kea97, BM88, Abe90, COW95, SAC99, MN99, CS04]. Researchers have observed that consumers are more likely to consume products that they have consumed in the past. There are two potential explanations for the sources of persistence in consumption choices. Heckman (1981) [Hec81] gives a clear definition of the two explanations:

One explanation is that as a consequence of experiencing an event, preferences, prices, or constraints relevant to future choices (or outcomes) are altered. In this case, past experience has a genuine behavioral effect in the sense that an otherwise identical individual who did not experience the event would behave differently in the future than an individual who experienced the event. ... A second explanation is that individuals may differ in certain unmeasured variables that influence their probability of experiencing the event but that are not influenced by the experience of the event.

Heckman refers to the first case as structural state dependence and the second case as unobserved heterogeneity. He also proposes a statistical model of discrete dynamic choice to distinguish the effects of these two concepts. Although his original work is applied to the U.S. labor market, Guadagni and Little (1983) [GL83] extend Heckman’s definitions to study observed persistence in consumption choices. Their 1983 [GL83] and 1998 [GL98] papers address the observed consumption persistence in the U.S. ground coffee market by using a multinomial logit choice model. Along this line, more recent works aim to distinguish structural state dependence and consumers' unobserved heterogeneity by using the individual-level purchase data. These works include [DM88, KR89, CJV91, JVC94, BG92, GC94, DHR10, BDG10]. For the application in consumption behavior, state dependence is referred to the causal link between past and present consumption behavior. Consumers’ unobserved heterogeneity is referred to as the consumers’ different exogenous preferences over brands that are unrelated to consumption history. Both patterns can explain the observed persistence in consumers’ brand choices.

In this paper, I simultaneously consider state dependence and consumers’ unobserved heterogeneity and use individual-level demographics and purchase data to distinguish the two effects. I study the inertia in consumption choices in the U.S. coffee market by looking at consumers’ consumption decisions over eight major coffee brands in the United States: Folgers, Maxwell House, Starbucks, Folgers Coffeehouse Series, Chock Full O’Nuts, Millstone, Eight O’Clock, and Yuban Coffee. The data that I use comprise the 2004 and 2005 Nielsen Homescan Panel
data. They contain detailed purchase history and demographic information for a massive number of consumers, as well as the features of the U.S. coffee market. My estimation consists of three parts. In the first part, I do not assume consumers’ unobserved heterogeneity, and I estimate the spurious state dependence effect by solving the discrete choice model. In the second part, I then consider both structural state dependence and consumers’ unobserved heterogeneity and distinguish these two effects by semiparametrically applying the discrete mass point algorithm. Unlike in most other works that assume a fully parametric model (e.g., see [BG92, GC94]), I propose a more flexible semiparametric model that is robust to model misspecification. Finally, I explore the potential correlated errors and endogeneity problem and compare different methods that are applied in the literature to address these problems.

On the empirical side, this paper finds that structural state dependence and consumers’ unobserved heterogeneity are two major sources of the observed persistence in consumption choices. The correlated errors have a negligible effect on consumption inertia. Without considering consumers’ unobserved heterogeneity, I find a positive and significant state dependence effect - the spurious state dependence defined by Heckman (1981) [Hec81]. However, when considering both state dependence and consumers’ unobserved heterogeneity, the estimation results suggest two types of consumers in the U.S. ground coffee market. One type of consumers treats ground bean coffee as normal goods, and the other treats ground bean coffee as inferior goods. The structural state dependence is positive and significant for one type of consumers but negative for the other type.

The rest of the paper proceeds as follows. The discrete choice model with state dependence and consumers’ unobserved heterogeneity is described in Section II. The 2004 and 2005 Nielsen Homescan Panel data and the features of the U.S. coffee market are described in Section III. Section IV discusses the observed purchasing inertia and shows the statistics of switching between coffee brands. Empirical estimation results are shown in Section V. In Section VI, potential correlated errors and the endogeneity problem are considered and different methods of solving them are compared. The paper concludes in Section VII.

II. The Model

The existing literature has addressed the importance of past consumption on current consumption choices. Distinguishing state dependence and consumers’ unobserved heterogeneity are of fundamental importance to understanding consumption behavior and helping firms make pricing strategies. In Section II.A, I first describe a baseline model to estimate the spurious state dependence effect without assuming consumers’ unobserved heterogeneity. Combining the ideas of the baseline model and the discrete mass point method, I propose a semiparametric extension to identify both effects in Section II.B.
A. Without Assuming Consumers’ Unobserved Heterogeneity

In this subsection, I consider a market with \( J \) differentiated alternatives, indexed as \( j = 1, \ldots, J \). The outside option is denoted as \( j = 0 \), representing the option of not purchasing any products. There are \( I \) consumers in the market, indexed as \( i = 1, \ldots, I \). The purchase history of \( T_i \) purchases for each individual \( i \) is observed. Since the number of \( T_i \) purchases can vary among individuals, I use unbalanced panel data. To specify the utility function, I accept the conclusion of Dube, Hitsch and Rossi (2010) [DHR10] that past consumption can directly affect consumers’ current utility by changing consumers’ preferences over products. Specifically, the utility \( U_{ijt} \) of consumer \( i \) purchasing product \( j \) at time \( t \) is assumed to have the following structure:

\[
U_{ijt} = \delta_{ijt} + \epsilon_{ijt} = \beta_{ijt}Z_{ijt} + \alpha P_{ijt} + \xi_j + \epsilon_{ijt},
\]

where \( Z_{ijt} \) and \( P_{ijt} \) denote, respectively, the observed attributes and the price of product \( j \) that consumer \( i \) purchases at time \( t \). \( \beta_{ijt} \) is the specific preference of consumer \( i \) for product \( j \) at time \( t \). To model state dependence, \( \beta_{ijt} \) is assumed to be a function of individual demographics and past consumption of the product:

\[
\beta_{ijt} = \lambda + \gamma X_{it} + \mu d_{ij(t-1)} + \nu_{ijt}.
\]

Here, \( X_{it} \) is individual demographic variables and \( d_{ij(t-1)} \) is the individual’s past consumption of product \( j \). \( d_{ij(t-1)} = 1 \) if consumer \( i \) purchases product \( j \) at time \( (t-1) \) and \( d_{ij(t-1)} = 0 \) otherwise. \( \nu_{ijt} \) is the i.i.d. noise term that follows the standard normal distribution. The parameter \( \mu \) measures the state dependence effect. The coefficient \( \gamma \) measures the income effect. The value of \( \alpha \) in (1) measures the price sensitivity and \( \xi_j \) is the product specific fixed effect, which contains the unobserved attributes of product \( j \). The shock term \( \epsilon_{ijt} \) contains the measurement error and “matching” between the individual \( i \) and the product \( j \) at the current time \( t \). As per convention, \( \epsilon_{ijt} \) is assumed to follow a type one extreme value distribution and i.i.d across individuals, products and times. There is also a default utility for the outside option:

\[
U_{i0t} = \epsilon_{i0t}, \quad \text{i.e.} \quad \delta_{i0t} = 0.
\]

Without assuming consumers’ unobserved heterogeneity in this subsection, \( \lambda, \gamma, \mu, \alpha \) and \( \xi_j \) are invariant cross individuals. In Section II.B, I shall relax this assumption and allow those parameters to be individual specific when consumers’ unobserved heterogeneity is taken into account.

With the utility (1) and (3), a consumer \( i \) makes a purchase decision by comparing the utility of the \( J \) products and the outside option. The consumer chooses product \( j \) if and only if \( j \) returns the highest utility, i.e.

\[
U_{ijt} \geq U_{ikt}, \quad \text{for } k = 0, 1, \ldots, J.
\]
Since the idiosyncratic component $\epsilon_{ijt}$ is i.i.d. with the type one extreme value distribution, the probability $Q_{ijt}$ of consumer $i$ choosing the product $j = 0, 1, \cdots, J$ at time $t$ has a closed form solution:

$$Q_{ijt} = \frac{\exp(\delta_{ijt})}{1 + \sum_{k=1}^{J} \exp(\delta_{ikt})}.$$  

Let $[J] = \{0, 1, \cdots, J\}$. In light of the consumption choices of consumer $i$ at time $t = 1, \ldots, T_i$, the individual likelihood function can be written as:

$$L_i = \prod_{t=1}^{T_i} \prod_{j \in [J]} Q_{ijt}^{I(E_{ijt})},$$

where $E_{ijt}$ is the event that consumer $i$ chooses the product $j$ at time $t$ and $I(E)$ is the indicator function of an event $E$. Assuming the independence among consumers, the aggregated log-likelihood (ALL) function is:

$$\text{ALL} = \sum_{i=1}^{I} \log(L_i) = \sum_{i=1}^{I} \sum_{t=1}^{T_i} \sum_{j \in [J]} \log Q_{ijt}^{I(E_{ijt})}.$$

Then the parameters are estimated by maximizing (7) and the results of the U.S. coffee market are shown in Section V.A.

### B. With State Dependence and Consumers’ Unobserved Heterogeneity

In this subsection, I propose a model to simultaneously capture the effects of state dependence and consumers’ unobserved heterogeneity. The two effects are distinguished by using the discrete mass point algorithm with the Bayesian Information Criterion (BIC). The discrete mass point algorithm is an influential method in the 1980s and 1990s. The key to the model used in this paper is the representation of consumers’ choice probabilities through a distribution in a number of heterogeneous market segments. The major assumption is that there is a fixed and finite number of segments in the market. Households belong to each of these segments with a certain probability. Then a household is assigned to the segment for which it has the largest posterior probability of membership.

Under this basic assumption, Dayton and Macready (1988) introduce a type of latent class model in which the probability of latent class membership depends on certain demographics and consumer specific characteristic variables [DM88]. Kamakura and Russell (1989) propose a multinomial logit-mixture model in which the probability distribution of the segment membership is invariant across individuals [KR89]. Gupta and Chintagunta (1994) further extend this model by allowing household specific probabilities of segment membership [GC94], which is the closest model to the one explored in this paper. In [GC94], the probability
\( \phi_{is} \) for a household \( i \) belonging to a segment \( s \) is assumed to take the following form:

\[
\phi_{is} = \frac{\exp(\bar{\alpha}_s + \bar{\gamma}_s D_i)}{\sum_{r=1}^{S} \exp(\bar{\alpha}_r + \bar{\gamma}_r D_i)},
\]

where \( \bar{\alpha}_s \) and \( \bar{\gamma}_s \) are unknown segment specific parameters to be estimated and denote the contribution of the various demographic variables to the probability of segment membership. A similar model with applications to the liquid laundry detergent market can be found in Bucklin and Gupta (1992) [BG92].

The consumer heterogeneity model (8) adopted in [GC94] has several drawbacks. First, the model is fully parametric and restrictive in the sense that the segment membership probability \( \phi_{is} \) has to depend on the demographic variables in the specified functional form defined in (8). This may cause a misspecification problem when we have little domain knowledge about the consumers segmentation information. Second, it is unclear which demographic variables affect the membership probabilities, which may further complicate the validity of the model.

Motivated by these concerns, I propose to estimate the probability distribution \( \phi_{is} \) in a nonparametric fashion and integrate the estimate into the state dependence model described in Section II.B. Therefore, I arrive at the semiparametric framework introduced below. Unlike the model of Gupta and Chintagunta (1994) [GC94], this method assumes minimal structural conditions on the segment membership probability distribution. Therefore, it is more flexible and robust for model misspecification. On the down side, the flexibility of this method detracts from its estimation efficiency and computational feasibility. In particular, I need a larger amount of panel data in the time dimension. Nonetheless, the feasibility depends on the quality and quantity of the specific dataset. In this paper, I use the Nielsen Homescan Panel data, which usually trace an individual for a long period of time. In view of the relatively high frequency of coffee purchases, this data provide the chance of having a large purchase history for each individual. Given that, directly identifying the individual segment membership probability is feasible, as will be shown in Section V. I emphasize here that the trade off between model misspecification and estimation efficiency and feasibility is problem dependent.

For notation convenience, I shall denote the set of parameters as

\[
\theta_s = \{\lambda_s, \gamma_s, \mu_s, \alpha_s, \xi_{0,s}, \xi_{1,s}, \ldots, \xi_{J,s}\}.
\]

The question now becomes how to estimate the distribution for \( \theta_s \). Under the main assumption, I can apply the discrete mass point method, which assumes a fixed number of segments for consumer heterogeneity. That is, \( \theta_s \) can take a finite number of values, \( \{\theta_1, \theta_2, \ldots, \theta_S\} \), with each representing a certain consumer type. For each consumer \( i \), there is a corresponding probability distribution \( \{\phi_{i1}, \phi_{i2}, \ldots, \phi_{iS}\} \) that represents the probability of consumer \( i \) in each
segment/consumer type and \( \sum_{s=1}^{S} \phi_{is} = 1 \).

For the moment, the number of mass points \( S \) is assumed to be known. Shortly, I shall discuss the selection of the optimal \( S \) by using the BIC. Given \( S \), let \( \{\theta_1, \theta_2, ..., \theta_S\} \) be the values of those mass points. For consumer \( i \), the probability of being in segment \( s \in \{1, \ldots, S\} \) is denoted by \( \phi_{is} \). Conditioning on being in a segment \( s \), consumer \( i \) will make a purchase decision by comparing the utility of all \( J \) products and the outside option. Let \( Q_{ijt,s} \) be the conditional probability of consumer \( i \) choosing product \( j \) on \( \theta_s \). Then, assuming the type one extreme value distribution for \( \epsilon_{ijt,s} \), similarly as (5), \( Q_{ijt,s} \) has a closed form solution given by

\[
Q_{ijt,s} = \frac{\exp(\delta_{ijt,s})}{1 + \sum_{k=1}^{J} \exp(\delta_{ikt,s})}, \quad j = 0, 1, \ldots, J,
\]

where

\[
\delta_{ikt,s} = \beta_{ijt,s} Z_{ijt} + \alpha_s \epsilon_{ijt} + \xi_{ijt,s} + \epsilon_{ijt,s},
\]
\[
\beta_{ijt,s} = \lambda_s + \gamma_s X_{it} + \mu_s d_{ij(t-1)} + \nu_{ijt,s}.
\]

Here, the errors \( \epsilon_{ijt,s} \) and \( \nu_{ijt,s} \) are also i.i.d. across segments. Given the event \( E_{ijt} \) (which has the same definition as in Section II.A), the conditional version of the individual likelihood function (6) of consumer \( i \) belonging to a segment \( s \) is:

\[
L_{i,s} = \prod_{t=1}^{T_i} \prod_{j \in [J]} Q_{ijt,s}^{I(E_{ijt})}.
\]

Now, the unconditional individual likelihood function for consumer \( i \) is obtained by the \( \{\phi_{is}\} \)-weighted average of (11) using the law of total probability:

\[
L_i = \sum_{s=1}^{S} \phi_{is} L_{i,s}.
\]

Finally, the ALL function, summing up all consumers, is easily seen as follows:

\[
ALL = \sum_{i \in I} \log(L_i) = \sum_{i \in I} \log \left[ \sum_{s=1}^{S} \left( \phi_{is} \prod_{t=1}^{T_i} \prod_{j \in [J]} Q_{ijt,s}^{I(E_{ijt})} \right) \right].
\]

Comparing (13) with (7), we see that the ALL (7) is a special case of (13) when \( S = 1 \), i.e. when there is no consumers’ unobserved heterogeneity in the market. Therefore, the proposed model with the simultaneous effects of state dependence and consumer heterogeneity has a more flexible structure and is less restrictive. The following two effects can be effectively investigated: (a) the values of the \( S \) mass points \( \Theta_S = \{\theta_1, \theta_2, ..., \theta_S\} \) can be estimated by using the discrete mass
It is assumed until this point that \( S \) is known in the model (10)–(13). In view of the number of parameters to be estimated, in particular the dependence on \( S \), determining the optimal number of mass points \( S \) is a highly non-trivial and important issue in the current framework. A small \( S \) may underfit the parameters in (13) and therefore underestimate the consumers’ heterogeneity effect, whereas a large \( S \) may introduce some artificial noise into the final model and render the output model hard to interpret or even incorrect. In light of this challenge, Bayesian Information Criterion (BIC) [Sch78] is used as the model selection criterion because BIC has a certain nice asymptotic model selection consistent property. The BIC is defined as:

\[
BIC(S) = -\text{ALL} + \frac{M_S}{2} \log(N),
\]

where \( \text{ALL} \) is the aggregated log-likelihood (13), \( M_S \) is the number of free parameters to be estimated, and \( N \) is the total number of observations in the data. Note that \( M_S \) increases in \( S \). Clearly, (14) balances between the fitness of data and the model complexity as measured by \( M_S \). Larger models are penalized more than smaller ones. The optimal number of the mass points \( S \) is thus defined as the one that minimizes the BIC value (14). The estimation results are shown in the later Section V.B.

III. Data Description and the U.S. Coffee Market

A. The Data

This paper uses Nielsen Homescan Panel data provided by the Kilts Data Center at the University of Chicago Booth School of Business. The data contain a longitudinal panel of approximately 40,000 to 60,000 U.S. households that continually provide information to Nielsen about their household demographics, what products they buy, how much the products cost, and when and where they make purchases. The data cover purchase information over a variety of food and non-food items across all retail outlets in all U.S. markets. In this paper, I study purchase behavior in the U.S. ground coffee market. I use Nielsen 2004 and 2005 Homescan Panel data to form an unbalanced panel of consumers’ ground coffee purchases. I control for consumers with a household size equal to 1, and I choose demographics \( X_{it} \) to represent consumer income levels. The study includes eight major coffee brands in the United States: Folgers, Maxwell House, Starbucks, Folgers Coffeehouse Series, Chock Full O’Nut, Millstone, Eight O’Clock Coffee, and Yuban Coffee. I choose the product attributes \( Z_{ijt} \) to be coffee sizes and the prices \( P_{ijt} \) to be prices per ounce. Income levels \( X_{it} \), product attributes \( Z_{ijt} \)

\(^1\)In the sequel, I shall abbreviate the eight brands as: Folgers, Maxwell, Starbucks, Folgers Series (or Series), Chock, Millstone, Eight, and Yuban, respectively.
and prices $P_{jt}$ are normalized by the standard deviations to obtain consistent scale. The goal is to estimate the model with individual level demographics and purchase data, as well as the features of the U.S. coffee market.

B. The U.S. Coffee Market

Coffee is the most popular drink in the United States. Coffee Statistics shows that among coffee drinkers, average consumption in the United States is 3.1 cups of coffee per day. More than half of the population (equivalent to 150 million Americans) drink coffee every day, and the number is growing over time. In this study, the eight most popular coffee brands mentioned in Section III.A are considered to investigate consumers’ consumption behavior.

Taking samples from Nielsen Homescan Panel data, I calculate the aggregate coffee purchases (i.e. market shares), the average coffee prices and sizes for these eight brands. Figure 1 shows the market shares calculated using the samples from the 2004 and 2005 Nielsen Homescan Panel data. From the figure, Maxwell House and Folgers have the largest market shares, with both exceeding 25%. Yuban coffee has the smallest market share with less than 5%. The other five brands have market shares ranging from 5% to 10% - from largest to smallest, Folgers Coffeehouse Series, Eight O’Clock, Starbucks, Chock Full O’Nuts, and Millstone.

I then calculate the average prices and sizes of the eight coffee brands and the results are shown in TABLE 1 and TABLE 2 below. Note that the average price here indicates price per unit. Later on in the estimation, I shall use price per ounce.

<table>
<thead>
<tr>
<th>Chock</th>
<th>Eight</th>
<th>Folgers</th>
<th>Maxwell</th>
<th>Yuban</th>
<th>Millstone</th>
<th>Starbucks</th>
<th>Series</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.48</td>
<td>4.36</td>
<td>4.72</td>
<td>3.81</td>
<td>4.43</td>
<td>5.22</td>
<td>8.34</td>
<td>4.87</td>
</tr>
</tbody>
</table>


Table 2—Average Sizes of Coffee Brands, Unit Ounce.

<table>
<thead>
<tr>
<th>Chock</th>
<th>Eight</th>
<th>Folgers</th>
<th>Maxwell</th>
<th>Yuban</th>
<th>Millstone</th>
<th>Starbucks</th>
<th>Series</th>
</tr>
</thead>
<tbody>
<tr>
<td>17.72</td>
<td>14.76</td>
<td>25.79</td>
<td>20.02</td>
<td>23.76</td>
<td>11.09</td>
<td>13.26</td>
<td>24.87</td>
</tr>
</tbody>
</table>


TABLE 1 shows that Starbucks is the most expensive brand, while Chock Full O’Nuts is the least expensive brand. The other six brands have average prices of around $5 - from most to least expensive, Millstone, Folgers Coffeehouse Series,
Folgers, Yuban, Eight O’Clock and Maxwell House. TABLE 2 shows the average sizes of these brands, which allows for a comparison based on price per ounce. The price per ounce for Chock Full O’Nuts, Eight O’Clock, Folgers, Maxwell House, Yuban, Millstone, Starbucks and Folgers Coffeehouse Series are $0.196, $0.295, $0.183, $0.190, $0.186, $0.471, $0.629, and $0.196 respectively. Starbucks has the highest per ounce price, and Millstone has the second highest. The other six brands have similar price per ounce. Among them, Folgers has the lowest price per ounce, and Yuban has the second lowest.

IV. Inertia in Coffee Purchases and Switching between Coffee Brands

In this section, I examine inertia in consumers’ coffee purchases. To address this topic, I first calculate the probabilities of consumers not purchasing, switching between purchasing and not purchasing, consuming a single brand, and the probability of switching between two, three, or more brands by using the ob-
served data. The results are shown in TABLE 3 below. It is clearly shown that less than 2% of all coffee consumers switch between more than two coffee brands within a two-year period. 61.55% of consumers show a strong inertia in their coffee purchases by not switching between coffee brands over time, and 34.09% of them switch between two coffee brands. TABLE 3 shows that consumption inertia commonly occurs in the U.S. ground coffee market. In addition, it shows that most people only switch within a very limited set of coffee brands. The results are consistent with the findings in [BDG10] and some others.

Table 3—Statistics of coffee brand switching.

<table>
<thead>
<tr>
<th>Percentage (%)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>no purchase</td>
<td>0.38</td>
</tr>
<tr>
<td>no purchase-buying</td>
<td>2.08</td>
</tr>
<tr>
<td>one brand</td>
<td>61.55</td>
</tr>
<tr>
<td>two brands</td>
<td>34.09</td>
</tr>
<tr>
<td>three brands</td>
<td>1.89</td>
</tr>
</tbody>
</table>


Next, the transition matrix is calculated to show the probabilities of switching between every two coffee brands in two consecutive periods. Recall that the purchase data comprise unbalanced panel data, which show a different number of purchase history data for consumers. The period here is thus defined as the date of purchase for each consumer rather than a specific, common time range. The transition matrix is shown in TABLE 4. The columns represent the previous period of coffee purchases and the rows represent the current period of coffee purchases. The entry \((i, j)\) is the probability of switching from \(j\) to \(i\).

From TABLE 4, it is clear that the transition matrix is diagonally dominated, meaning that people have a high probability of being persistent with their consumption choices over time. A closer inspection shows that Eight O’Clock, Maxwell House and Starbucks consumers exhibit the most persistence over time, with a probability 0.77-0.78 of staying with those brands. The table implies that once people switch into those coffee brands, they are very likely to remain with those brands for a long time. Folgers Coffeehouse Series has the least consumption persistence. Only 50% of the people consuming Folgers Coffeehouse Series in the previous period choose to stay with that brand. Overall, around one third of people switch their coffee choices over time. Among the switchers, most of them are switching into Maxwell House. Maxwell House, with the highest probability of making consumers stay and the highest probability of attracting new switchers, obtains the largest market share at no surprise.

The transition matrix in TABLE 4 emphasizes that consumers exhibit inertia in their consumption choices over time, which motivates this study of the reasons behind consumption inertia. There are three potential explanations for the
### Table 4—Transition probability matrix of coffee brand switching.

<table>
<thead>
<tr>
<th></th>
<th>Chock</th>
<th>Eight</th>
<th>Folges</th>
<th>Maxwell</th>
<th>Yuban</th>
<th>Millstone</th>
<th>Starbucks</th>
<th>Series</th>
<th>No†</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chock</td>
<td>0.60</td>
<td>0.03</td>
<td>0.01</td>
<td>0.04</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.04</td>
<td>0</td>
</tr>
<tr>
<td>Eight</td>
<td>0.02</td>
<td>0.78</td>
<td>0.02</td>
<td>0.03</td>
<td>0.02</td>
<td>0.03</td>
<td>0</td>
<td>0.15</td>
<td>0</td>
</tr>
<tr>
<td>Folges</td>
<td>0.02</td>
<td>0.03</td>
<td>0.68</td>
<td>0.09</td>
<td>0.04</td>
<td>0.13</td>
<td>0.03</td>
<td>0.26</td>
<td>0</td>
</tr>
<tr>
<td>Maxwell</td>
<td>0.28</td>
<td>0.11</td>
<td>0.17</td>
<td>0.78</td>
<td>0.08</td>
<td>0.02</td>
<td>0.07</td>
<td>0.14</td>
<td>0.15</td>
</tr>
<tr>
<td>Yuban</td>
<td>0.04</td>
<td>0.01</td>
<td>0.02</td>
<td>0.01</td>
<td>0.72</td>
<td>0</td>
<td>0.02</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Millstone</td>
<td>0</td>
<td>0.03</td>
<td>0.02</td>
<td>0.02</td>
<td>0.04</td>
<td>0.65</td>
<td>0.02</td>
<td>0.05</td>
<td>0.15</td>
</tr>
<tr>
<td>Starbucks</td>
<td>0.02</td>
<td>0</td>
<td>0.01</td>
<td>0.01</td>
<td>0.04</td>
<td>0.09</td>
<td>0.77</td>
<td>0.01</td>
<td>0.15</td>
</tr>
<tr>
<td>Series</td>
<td>0.02</td>
<td>0</td>
<td>0.06</td>
<td>0.03</td>
<td>0.04</td>
<td>0.07</td>
<td>0.02</td>
<td>0.50</td>
<td>0</td>
</tr>
<tr>
<td>No</td>
<td>0</td>
<td>0.01</td>
<td>0</td>
<td>0.04</td>
<td>0.02</td>
<td>0.05</td>
<td>0</td>
<td>0.39</td>
<td></td>
</tr>
</tbody>
</table>

| Total | 1       | 1       | 1       | 1       | 1     | 1         | 1         | 1      | 1   |

*“No” means “No Purchase”.

observed consumption inertia I shall discuss below.

1) First, past consumption could directly influence the consumer choices in the current period, which is known as the state dependence effect. Consumers choose a certain coffee brand in the current period because they consumed the same brand in the previous period. In this case, simply shuffling consumers' initial choices would change the probability of their later consumption choices and thus the market shares of those brands.

2) Second, consumers differ along some serially correlated unobserved propensity to make consumption decisions, which is called consumers’ unobserved heterogeneity. In this case, consumers choose a certain brand not because they consumed the same brand before, but rather because they like the brand better. By contrast, reshuffling their initial coffee brand choices does not change the market pattern much.

3) Third, observed inertia could come from the potential correlated error terms. People do not consume a product today because they consumed the same product in the last period, or because they like the product better than others; rather they face some idiosyncratic shocks that tend to be persistent over time.
To summarize, testing the existence of those potential sources of consumption inertia and distinguishing their effects are of fundamental importance to marketing and firm price setting. In the next section, I consider state dependence and consumers’ unobserved heterogeneity using the model in Section II.B. Estimation results for the Nielson Homescan Panel data are shown in Section V. I also address the potential correlated error problem in Section VI.

V. Estimation Results

In this section, the estimation results for the two cases described in Section IV are presented. Estimation results using the model (cf. Section II.A) without assuming consumers’ unobserved heterogeneity are shown in Section V.A. Parameters (\(\lambda, \gamma, \mu, \alpha, \xi_{jt}\)) in this model are consumer invariant, and they are estimated by maximizing the aggregated log-likelihood (7). Then, in Section V.B, both the state dependence and consumers’ unobserved heterogeneity are considered. In this model (cf. Section II.B), the distribution \(\{\phi_{is}\}_{s=1}^{S}\) of \(\{\theta_{s}\}_{s=1}^{S}\), \(\theta_{s} = \{\lambda_{s}, \gamma_{s}, \mu_{s}, \alpha_{s}, \xi_{s0t}, \xi_{s1t}, \ldots, \xi_{sJt}\}\), is estimated by applying the discrete mass point algorithm and the optimal number of mass points \(S\) is determined by the BIC.

A. Without Assuming Consumers’ Unobserved Heterogeneity

Here, consumers’ unobserved heterogeneity is not included in the model. The estimation results for the model parameters in Section II.A are shown in TABLE 5 and TABLE 6. TABLE 5 shows the estimated constant term \(\lambda\), income effect \(\gamma\), spurious state dependence effect \(\mu\), and price sensitivity \(\alpha\). From the table, the past consumption effect, i.e the spurious state dependence, is positive and significant. The significant positive past consumption effect implies that for a given brand, consuming it during the last period would dramatically increase the chance of consuming it during the current period. This estimation is consistent with the brand switching statistics of the transition matrix in TABLE 4. The income effect is positive, implying that having a higher income would increase the consumption of ground coffee. Thus, ground coffee is a normal good for the average consumer. The price effect is negative, meaning that a higher price would discourage consumption of ground coffee.

Table 5—Estimated Coefficients without Assuming Consumers’ Unobserved Heterogeneity.

\[
\begin{array}{cccc}
\lambda & \gamma & \mu & \alpha \\
-0.5235 & 0.1230 & 0.9792 & -0.7304 \\
(0.0099) & (0.0002) & (0.0682) & (0.0035) \\
\end{array}
\]

\(\dagger\)\(\lambda\) is the constant term, \(\gamma\) is the income effect, \(\mu\) is the spurious state dependence effect and \(\alpha\) is the price sensitivity.
TABLE 6 shows the estimated coffee brand fixed effects without consumer heterogeneity. The fixed effects are normalized such that the effect of Folgers Coffeehouse Series is zero. From the table, Folgers has the highest brand fixed effect. Maxwell House seconds with a slightly lower brand fixed effect. Yuban coffee has the lowest brand fixed effect, and Millstone coffee has the second lowest. These results are highly correlated with the observed market share pattern in FIGURE 1. In other words, Maxwell House and Folgers have the highest brand fixed effects while they have the highest observed market shares. Yuban and Millstone coffee have the lowest brand fixed effects while they have the smallest market shares. For the rest four brands, the order from highest to lowest fixed effect is Starbucks, Chock Full O’Nuts, Folgers Coffeehouse Series, and Eight O’Clock.

Table 6—Estimated coffee brand fixed effects, without assuming consumers’ unobserved heterogeneity.

<table>
<thead>
<tr>
<th></th>
<th>Chock</th>
<th>Eight</th>
<th>Folgers</th>
<th>Maxwell</th>
<th>Yuban</th>
<th>Millstone</th>
<th>Starbucks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimate</td>
<td>0.5396</td>
<td>-0.3780</td>
<td>2.9051</td>
<td>1.7273</td>
<td>-2.6620</td>
<td>-1.4814</td>
<td>0.7472</td>
</tr>
<tr>
<td>Standard Error</td>
<td>(0.0028)</td>
<td>(0.0016)</td>
<td>(0.0304)</td>
<td>(0.0041)</td>
<td>(0.0231)</td>
<td>(0.0468)</td>
<td>(0.0174)</td>
</tr>
</tbody>
</table>

Based on the estimation results, I then compute the probability of purchasing each coffee brand, given purchasing and not purchasing the same brand in the previous period, respectively. For each income level, I calculate the probability of purchasing each coffee brand conditional on \(d_{ij(t-1)} = 0\) and \(d_{ij(t-1)} = 1\). I then take the difference of these two values to show how much the probability increases by purchasing the same brand in the previous period. I adopt the estimated values of parameters from TABLE 3 and TABLE 4 and use the average prices and sizes for each coffee brand from TABLE 1 and TABLE 2. The calculated increases in probabilities across coffee brands and income levels are shown in TABLE 7.

From TABLE 7, it is clearly shown that for each coffee brand and income level, there is a probability increase when the same brand is purchased in the previous period. However, the magnitudes of probability gains vary across brands and income levels. Comparing across each row in TABLE 7, Maxwell House and Folgers have the strongest state dependence effects. Consumers who purchased these two brands in the last period have a 35% increase in probability of purchasing the same brands in the current period. Chock Full O’Nuts and Folgers Coffeehouse Series have slightly lower state dependence effects, with the probability gains around 14%. Starbucks, Eight O’Clock, Yuban, and Millstone coffee have the lowest levels of state dependence effects, with the probability gains less than 5%. A comparison across income levels shows that as incomes go up, the gain in probability increases for Yuban and Folgers Coffeehouse Series, roughly stays the same for Maxwell House, and decreases for Chock Full O’Nuts, Eight O’Clock, Folgers, Millstone, and Starbucks. The overall effect of income levels does not
Table 7—Increase in Consumption Probability with State Dependence Effect, across coffee brands and income levels.

<table>
<thead>
<tr>
<th>Income</th>
<th>Chock</th>
<th>Eight</th>
<th>Folgers</th>
<th>Maxwell</th>
<th>Yuban</th>
<th>Millstone</th>
<th>Starbucks</th>
<th>Series</th>
</tr>
</thead>
<tbody>
<tr>
<td>$5000 -</td>
<td>0.1464</td>
<td>0.0372</td>
<td>0.3649</td>
<td>0.3394</td>
<td>0.0113</td>
<td>0.0045</td>
<td>0.0259</td>
<td>0.1402</td>
</tr>
<tr>
<td>$5000</td>
<td>0.1455</td>
<td>0.0366</td>
<td>0.3579</td>
<td>0.3394</td>
<td>0.0115</td>
<td>0.0044</td>
<td>0.0253</td>
<td>0.1423</td>
</tr>
<tr>
<td>$8000</td>
<td>0.1445</td>
<td>0.0359</td>
<td>0.3511</td>
<td>0.3394</td>
<td>0.0116</td>
<td>0.0043</td>
<td>0.0247</td>
<td>0.1442</td>
</tr>
<tr>
<td>$10,000</td>
<td>0.1433</td>
<td>0.0352</td>
<td>0.3442</td>
<td>0.3391</td>
<td>0.0118</td>
<td>0.0041</td>
<td>0.0241</td>
<td>0.1461</td>
</tr>
<tr>
<td>$12,000</td>
<td>0.1427</td>
<td>0.0348</td>
<td>0.3409</td>
<td>0.3390</td>
<td>0.0118</td>
<td>0.0040</td>
<td>0.0238</td>
<td>0.1470</td>
</tr>
<tr>
<td>$15,000</td>
<td>0.1417</td>
<td>0.0341</td>
<td>0.3342</td>
<td>0.3386</td>
<td>0.0119</td>
<td>0.0039</td>
<td>0.0232</td>
<td>0.1487</td>
</tr>
<tr>
<td>$20,000</td>
<td>0.1401</td>
<td>0.0333</td>
<td>0.3277</td>
<td>0.3382</td>
<td>0.0121</td>
<td>0.0038</td>
<td>0.0226</td>
<td>0.1504</td>
</tr>
<tr>
<td>$25,000</td>
<td>0.1394</td>
<td>0.0330</td>
<td>0.3244</td>
<td>0.3379</td>
<td>0.0121</td>
<td>0.0037</td>
<td>0.0223</td>
<td>0.1512</td>
</tr>
<tr>
<td>$30,000</td>
<td>0.1387</td>
<td>0.0326</td>
<td>0.3212</td>
<td>0.3376</td>
<td>0.0122</td>
<td>0.0036</td>
<td>0.0219</td>
<td>0.1519</td>
</tr>
<tr>
<td>$40,000</td>
<td>0.1379</td>
<td>0.0323</td>
<td>0.3181</td>
<td>0.3373</td>
<td>0.0122</td>
<td>0.0036</td>
<td>0.0216</td>
<td>0.1527</td>
</tr>
<tr>
<td>$45,000</td>
<td>0.1372</td>
<td>0.0319</td>
<td>0.3149</td>
<td>0.3370</td>
<td>0.0123</td>
<td>0.0035</td>
<td>0.0213</td>
<td>0.1534</td>
</tr>
<tr>
<td>$50,000</td>
<td>0.1356</td>
<td>0.0311</td>
<td>0.3088</td>
<td>0.3362</td>
<td>0.0123</td>
<td>0.0034</td>
<td>0.0207</td>
<td>0.1548</td>
</tr>
<tr>
<td>$60,000</td>
<td>0.1340</td>
<td>0.0304</td>
<td>0.3027</td>
<td>0.3354</td>
<td>0.0124</td>
<td>0.0032</td>
<td>0.0201</td>
<td>0.1562</td>
</tr>
<tr>
<td>$70,000</td>
<td>0.1314</td>
<td>0.0293</td>
<td>0.2939</td>
<td>0.3340</td>
<td>0.0125</td>
<td>0.0031</td>
<td>0.0192</td>
<td>0.1580</td>
</tr>
<tr>
<td>$100,000+</td>
<td>0.1306</td>
<td>0.0289</td>
<td>0.2911</td>
<td>0.3335</td>
<td>0.0126</td>
<td>0.0030</td>
<td>0.0189</td>
<td>0.1586</td>
</tr>
</tbody>
</table>
indicate a strong trend in the gain in probability.

To summarize, the estimation results imply a positive and significant past consumption effect. Consuming the product in the last period significantly increases the chance of consuming the same product in the current period. The magnitudes of probability increases vary across coffee brands, with Maxwell House and Folgers exhibiting the strongest magnitudes. On the contrary, income level does not have a significant effect on the probability gains. Nevertheless, the average income effect is positive, which implies that consumers in general treat ground coffee as normal goods. In the next subsection, I shall discuss the potential explanations for the positive spurious state dependence effect.

B. Consumers’ Unobserved Heterogeneity, Discrete Mass Point Algorithm

In the previous subsection, I explain the estimation of the discrete choice model without assuming consumers’ unobserved heterogeneity. The estimation results show a positive and significant spurious state dependence effect. This positive effect could be explained by two factors: state dependence and consumers’ unobserved heterogeneity. The state dependence refers to the causal links between past and present purchase behavior, while consumers’ unobserved heterogeneity refers to consumers’ different exogenous preferences over brands that are unrelated to the consumers’ past purchase history. In this subsection, consumers’ unobserved heterogeneity is considered in both their responses (slope) and preferences (intercept). The model selection and parameter estimation are referred to in Section II.B.

The optimal model selected by the BIC contains two types of consumers in the U.S. ground coffee market, i.e $S = 2$. These two types of consumers are referred to as S1 and S2 in the sequel. The estimated parameters for both types of consumers are shown in TABLE 8 and TABLE 9.

Table 8—Estimated coefficients for the two types of consumers, by assuming consumers’ unobserved heterogeneity.

<table>
<thead>
<tr>
<th></th>
<th>λ</th>
<th>γ</th>
<th>μ</th>
<th>α</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>-0.2830</td>
<td>0.4483</td>
<td>1.0928</td>
<td>-0.6415</td>
</tr>
<tr>
<td></td>
<td>(0.0004)</td>
<td>(0.0039)</td>
<td>(0.0334)</td>
<td>(0.0073)</td>
</tr>
<tr>
<td>S2</td>
<td>-0.0103</td>
<td>-0.4663</td>
<td>-0.5108</td>
<td>1.5711</td>
</tr>
<tr>
<td></td>
<td>(0.0000)</td>
<td>(0.0012)</td>
<td>(0.0057)</td>
<td>(0.0253)</td>
</tr>
</tbody>
</table>

†λ is the constant term, γ is the income effect, μ is the structural state dependence effect and α is the price sensitivity.

TABLE 8 shows that S1 consumers have a positive income effect, a positive and significant state dependence effect, and a negative price elasticity. S2 consumers by contrast have a negative income effect, a negative state dependence effect, and
a positive price elasticity. For S1 consumers, the positive income effect implies
that a higher income would increase the probability of consuming each coffee
brand. Thus ground coffee is a normal good for type one consumers. For S2
consumers, the income effect $\gamma$ is negative. The potential explanation for the
negative income effect is that for a given product with attributes $Z_{jt}$ and price
$P_{jt}$, a higher income would lower consumption of that product. Essentially, people
with a higher income tend to buy hot coffee from coffee stores instead of buying
ground coffee and making it at home. Thus, for S2 consumers, the ground coffee
is an inferior good. Moreover, S1 consumers show a strong persistence in their
coffee brand choices by having a positive and significant state dependence effect.
S2 consumers, on the other hand, do not show persistence in their choices of coffee
brands over time and have a negative state dependence coefficient.

Researchers have given several explanations for the positive and negative state
dependence effects. A popular explanation for the positive state dependence is
that past consumption could cause a brand loyalty through a form of psychological
switching cost or a habit persistence mechanism [GL83, Kea97, FK07, DHR10].
Alternative explanation for the positive structural state dependence is consumer
learning behavior [Osb07, MS02, HN06, EKS08, CEK13]. Consumers collect in-
formation about products during consumption and more consumption in the past
implies less amount of learning needed to know the product quality. Thus, con-
sumers would like to consume products they have experienced before in order to
avoid the new learning process. Another similar argument is consumers' searching
costs. Consumers face searching costs when making a purchase in a product
category and thus do not consider brands that they have not recently bought.
Dube, Hitsch and Rossi (2010) [DHR10] explore all the explanations above and
conclude that the form of structural state dependence shown in the data is con-
sistent with loyalty, but not with searching or learning. By contrast, the Erdem
and Keane (1996) model and the large body of subsequent work derived from it
take stand that state dependence derives from the learning mechanism.

The negative state dependence effect and the potential explanations are doc-
umented in the literature as well. One potential explanation is that consumers
are variety-lovers (documented in [Kea97, Pau11]) and hence like to try different
coffee brands. Overall, which explanations make the most sense remains an open
research question and I shall leave it for future study.

Besides the explanations of state dependence, TABLE 8 also implies that for S1
consumers, a higher price decreases their chance of consuming a certain brand,
while it is the opposite for S2 consumers. For S2 consumers, a higher price
increases their chance of consuming a certain brand. This result is consistent with
my explanation that S1 consumers treat ground coffee as normal goods while S2
consumers treat ground coffee as inferior goods. Therefore, for S1 consumers, they
consider prices a lot when making consumption decisions. Thus a higher price
would discourage them from consuming an expensive brand. For S2 consumers,
you may choose a better-quality coffee brand with a higher price. Overall, the
two types of consumers are opposite in their consumption behavior because they have different preferences in income, state dependence and price.

**Table 9—Coffee Brand Fixed Effects, by assuming Consumers’ Unobserved Heterogeneity**

<table>
<thead>
<tr>
<th></th>
<th>Chock</th>
<th>Eight</th>
<th>Folgers</th>
<th>Maxwell</th>
<th>Yuban</th>
<th>Millstone</th>
<th>Starbucks</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>-0.5164</td>
<td>-0.7101</td>
<td>3.1034</td>
<td>2.067</td>
<td>-2.2952</td>
<td>-0.9432</td>
<td>1.5626</td>
</tr>
<tr>
<td></td>
<td>(0.0041)</td>
<td>(0.0098)</td>
<td>(0.1148)</td>
<td>(0.1862)</td>
<td>(0.2020)</td>
<td>(0.0132)</td>
<td>(0.0470)</td>
</tr>
<tr>
<td>S2</td>
<td>-0.1748</td>
<td>-0.1547</td>
<td>-2.7114</td>
<td>-1.3349</td>
<td>2.2828</td>
<td>1.9084</td>
<td>-2.1553</td>
</tr>
<tr>
<td></td>
<td>(0.0003)</td>
<td>(0.0008)</td>
<td>(0.0024)</td>
<td>(0.0271)</td>
<td>(0.0430)</td>
<td>(0.0325)</td>
<td>(0.0366)</td>
</tr>
</tbody>
</table>

TABLE 9 shows the preferences for coffee brands for both types of consumers. S1 consumers have the highest brand preferences for Folgers and Maxwell House and the lowest brand preferences for Yuban and Millstone coffee. S2 consumers have the highest brand preferences for Yuban and Millstone coffee and have the least preference for Folgers and Starbucks coffee. Thus, these two types of consumers have opposite preferences for coffee brands.

To conclude this section, there are two types of consumers existing in the U.S. ground coffee market, and they have very different consumption preferences in the following aspects:

1) S1 consumers have a positive income effect and a negative price sensitivity. S2 consumers by contrast have a negative income effect and a positive price sensitivity. To further investigate, S1 consumers treat ground coffee as normal goods while S2 consumers treat ground coffee as inferior goods.

2) S1 consumers have a significant and positive state dependence effect, which explains the strong persistence in their consumption choices over time. S2 consumers have a negative state dependence effect possibly because they love variety. The results are consistent with the observations in the transition matrix in TABLE 4.

3) The two types of consumers have opposite preferences for coffee brands. S1 consumers prefer Folgers and Maxwell House the most and Yuban and Millstone coffee the least. S2 consumers, to the opposite, prefer Yuban and Millstone coffee the most and Folgers and Starbucks coffee the least.

VI. Potential Correlated Errors and Further Challenges

**A. Potential Correlated Errors**

In the previous section, the model is estimated with the assumption that the error terms $e_{ijt}$ in the utility function (1) are i.i.d across individuals, products
and times. Recent work by Keane (1997) [Kea97] argues that it is reasonable to allow error terms to be correlated. Other relevant studies include [Cha85, Kea94, EK96, HK00, Nes02, DHRV08, DHR10]. The possible correlation of error terms may introduce a bias to the estimation results since the past consumption $d_{ij(t-1)}$ is a function of the past error term $\epsilon_{ij(t-1)}$ and hence it is correlated with the current error term $\epsilon_{ijt}$. Thus, a past consumption could account for a large current random utility draw and would predict the current choice behavior.

Given the potential correlated errors and combining it with the previous analysis, persistence in consumption choices may arise from three sources: (a) structural state dependence, (b) consumers’ unobserved heterogeneity, and (c) serial correlation in the error terms. Realizing and identifying the three potential sources of consumption inertia is not an easy empirical task. The previous sections present an analysis of the first two sources. This section will be devoted to explore the potential correlated errors concern.

Several approaches from recent literature are applied to solve the correlated errors problem. Chamberlain (1985) [Cha85] proposes to use lagged prices as proxy for state dependence term. Different from my approach, he adopts a five-component normal mixture for heterogeneity model. Keane (1997) [Kea97] uses simulation method to consider both state dependence and correlated errors, though he assumes a specific AR(1) process of the error terms. Dube, Hitsch and Rossi (2010) [DHR10] argue that the correlation between the past consumption state and the current product choice should be lower if the loyalty state was initiated by a price discount rather than a regular price. They examine the consumption inertia by including a discount indicator in the utility form and check if the coefficient of the discount indicator is negative or zero. Their work also assumes the specific AR(1) process of the error terms.

In this section, I examine the potential correlated errors by using the past prices as proxy for the past consumption indicator $d_{ij(t-1)}$ for the following reasons. First, the lack of discount data disallows me to include a discount indicator in the utility form. Second, I try to approach the question by imposing the minimum structural assumption on the error terms (such as the AR(1) process) in order to avoid the model misspecification. Therefore, (2) in the state dependence model now becomes:

\begin{equation}
\beta_{ijt} = \lambda + \gamma X_{it} + \mu P_{ij(t-1)} + \nu_{ijt}.
\end{equation}

It is generally agreed that the lagged prices go directly into the lagged utility form and hence are correlated with the past consumption indicator $d_{ij(t-1)}$. Also the past prices $P_{ij(t-1)}$ do not influence the persistence in consumption choices over time and thus are uncorrelated with the error term $\epsilon_{ijt}$. Therefore, the past prices are considered a valid proxy for state dependence term. Using past prices as proxy for past consumption, I repeat the analysis in Section V and the estimation results are shown in TABLE 10 and TABLE 11 below.
Table 10—Estimated coefficients using past price as proxy for past consumption.

<table>
<thead>
<tr>
<th></th>
<th>$\lambda$</th>
<th>$\gamma$</th>
<th>$\mu$</th>
<th>$\alpha$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-0.5118</td>
<td>0.0720</td>
<td>1.0950</td>
<td>-0.7270</td>
</tr>
<tr>
<td></td>
<td>(0.0042)</td>
<td>(0.0002)</td>
<td>(0.0095)</td>
<td>(0.0024)</td>
</tr>
</tbody>
</table>

$\dagger$ $\lambda$ is the constant term, $\gamma$ is the income effect, $\mu$ is the spurious state dependence effect and $\alpha$ is the price sensitivity.

Table 11—Estimated coffee brand fixed effects, using past price as proxy for past consumption.

<table>
<thead>
<tr>
<th>Brand</th>
<th>Chock</th>
<th>Eight</th>
<th>Folgers</th>
<th>Maxwell</th>
<th>Yuban</th>
<th>Millstone</th>
<th>Starbucks</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.5566</td>
<td>-0.3680</td>
<td>2.9078</td>
<td>1.7476</td>
<td>-2.6714</td>
<td>-1.4491</td>
<td>0.7603</td>
</tr>
<tr>
<td></td>
<td>(0.0014)</td>
<td>(0.0025)</td>
<td>(0.0460)</td>
<td>(0.0937)</td>
<td>(0.1611)</td>
<td>(0.0015)</td>
<td>(0.0204)</td>
</tr>
</tbody>
</table>

Comparing Table 10 and Table 11 with Table 5 and Table 6, it is clearly shown that the estimation results do not change much by using past prices as proxy for state dependence term. Thus the correlated errors are found to contribute negligibly to the consumption persistence observed in the U.S. coffee market. As a caveat to interpret this result, although the results show the minimum effect of correlated errors, those numbers could not lead further quantifying analysis of the amount of each effect.

Findings from other researchers reinforce my conclusion. Chamberlain [Cha85] shows that the correlated errors do not greatly affect the estimation results, and Keane (1997) [Kea97] concludes that the estimated degree of state dependence remains highly unchanged if autocorrelated random utility terms are allowed for. In addition, the estimation results by Dube, Hitshc and Rossi (2010) [DHR10] imply that overall there is a scant evidence that the measured state dependence is due to autocorrelated errors.

B. Further Challenges

[Endogeneity]. Although researchers have recently solved the correlated errors problem, some degree of the endogeneity problem persists. The past consumption indicator term $d_{ij(t-1)}$ confounds both structural state dependence and consumers’ unobserved heterogeneity in the previous period. Therefore, the past consumption indicator term is endogenous in the sense that it correlates with all the preference coefficients that shall be estimated. The methods that do not take into account this endogeneity problem could cause bias to the estimated parame-

$^2$Thanks to J. P. Dube for pointing out in class.
ters. To solve this problem, Paulson in his 2011 and 2012 [Pau11] [Pau12] papers apply the conditional covariates method to identify structural state dependence in a binary choice model. Paulson concludes that the commonly used methods, which do not take into account the endogeneity problem, overestimate the amount of structural state dependence. His method, however, requires a strong conditional exogeneity assumption and does not apply to the multinomial model.

Other researchers have proposed alternatives. Honore and Kyriazidou (2000) [HK00] propose a semiparametric identification strategy. Their estimator can disentangle structural state dependence but cannot handle the correlated errors. Arellano and Carrasco (2003) [AC03] presents a random effect identification for a binary choice panel model. Their method requires a specific distribution for the unobserved individual-specific time-varying components. Blundell and Powell (2004) [BP04] and Hoderlein (2008) [Hod08] both use a control function approach, however their method cannot apply to the lagged dependent variable because there is not a plausible instrument available for the lagged dependent variable. Overall, all of these methods require a strong assumption in order to work, and none of them can apply easily to the multinomial logit model, with both lagged consumption and the correlated error terms included. It remains an open research question to be considered in the future work.

[Availability of coffee brand]. Besides the endogeneity problem, the availability of coffee brands also raises concern. In the paper, it is assumed that all eight coffee brands are available per purchase per consumer. However, it is not likely to be true in reality. Consumers may face fewer choices when they make purchase decisions. It could be because the grocery stores they purchase coffee from have only limited coffee brands on the self. Therefore, it adds to one potential source of consumption inertia. People are persistent with their coffee brand choices over time or people are switching within a limited number of coffee brands could either due to the structural state dependence, consumers’ unobserved heterogeneity, or it could be because they do not have many choices to choose from. Thus, the availability of coffee brands plays an important role in understanding consumption persistence.

To address this issue, future work could be done to trace the brand availabilities of grocery stores where consumers make their purchases. Alternatively, and perhaps simpler, I may trace a certain type of grocery store chains in multiple locations across the country (e.g. CVS, Walgreen or Walmart). In that case, I would be able to control the availabilities of coffee brands per each consumer. In both ways, detailed data of grocery store supplies are required. Nielsen provides those data in a separate data file and I shall combine them together for further investigation on the brand availability problem.

VII. Conclusion

Understanding how consumption choices are persistent and distinguishing the factors that contribute to consumption inertia are of fundamental importance to
marketing and pricing policy making.

In this paper the observed persistence of consumption choices in the U.S. coffee market is examined. Nielsen Homescan Panel data, which contain individual level demographics, purchase information and the features of the U.S. coffee market are used. The paper forms the discrete choice model, with including past consumption in the utility form. In the first part, the effect of past consumption on current consumption choices (i.e. spurious state dependence) is estimated without assuming consumers’ unobserved heterogeneity. The estimation results show a positive and significant past consumption effect. Thus, consuming the product in the last period could significantly increase the chances of consuming the same product in the current period.

Next, structural state dependence and consumers’ unobserved heterogeneity are considered. The effects of these factors are explored by applying the discrete mass point algorithm along with the Bayesian Information Criterion (BIC). The estimation results suggest that there are two types of consumers on average existing in the U.S. ground coffee market. The first type of consumers has a positive income effect, a positive and significant state dependence effect and a negative price sensitivity. The second type of consumers by contrast has a negative income effect, a negative state dependence effect and a positive price sensitivity. The two types of consumers have very different consumption behavior. In addition, they have opposite preferences for coffee brands. The first type of consumers prefers Maxwell House and Folgers coffee the most and Yuban and Millstone coffee the least, while the second type of consumers prefers Yuban and Millstone coffee the most and Folgers and Starbucks coffee the least.

There are several potential explanations for the results. The first type of consumers has a positive income effect and treats ground coffee as normal goods, while the second type of consumers has a negative income effect and treats ground coffee as inferior goods. The first type of consumers consumes more ground coffee when their income increases, while the other type of consumers chooses to consume less ground coffee when their income goes up, potentially because they may switch to the hot coffee at the coffee stores instead of purchasing ground coffee and making it at home. The structural state dependence effect is positive and significant for the first type of consumers, but is negative for the other type. In the literature, there are potential explanations for both positive and negative state dependence effects. Possible explanations for a positive state dependence effect include brand loyalty, learning behavior and switching costs, which give people incentives to stay with their choices over time. The negative state dependence effect implies that people tend to diverge in their consumption choices. This could be because some people like variety and try different brands.

In addition, the paper considers potential correlated errors. The paper tests the effect of it on consumption persistence by using past product price as proxy for past consumption. The empirical results imply that correlated errors make negligible contribution to the consumption inertia. Overall, the paper finds that
both structural state dependence and consumers’ unobserved heterogeneity are the two main explanations of the consumption persistence observed in the U.S. coffee market. The findings could help explain firms’ marketing and price setting incentives. This paper contributes to the literature of understanding consumption persistence particularly in the U.S. coffee market.

REFERENCES


