# State Space Modelling of Dynamic Choice Behavior with Habit Persistence

A Dissertation Presented for the
Doctor of Philosophy
Degree
The University of Tennessee, Knoxville

Kang Bok Lee
August 2014

## Dedication

With gratitude, I dedicate this dissertation to my wife, Sumin Han, to my parents, Hoonkoo Lee and Myunghee Choi, to my sister Sora Lee, and to my other relatives and friends who have provided support and encouragement throughout my years in the PhD program.

## Acknowledgements

First and foremost I wish to thank my advisor, Professor Russell Zaretzki, for supporting me during these past four years. Russell has given me the freedom to pursue various projects without objection. I am also very grateful to him for his scientific advice and knowledge and many insightful discussions and suggestions. He is my primary resource for getting my statistics questions answered and one of the smartest people I know. I greatly value the close personal rapport that Russell and I have forged over the years. I quite simply cannot imagine a better adviser.

I would also like to thank my committee members, Professors Chad Autry, Christian Vossler, Randy Bradley, Bogdan Bichescu, and Neeraj Bharadwaj. I am deeply indebted to Professors Chad Autry and Christian Vossler for their important role in my doctoral work. I hope that I could be as lively, enthusiastic, and energetic as Chad and to someday be able to command an audience as well as he can. I would particularly like to acknowledge Christian, who has long been an inspiring figure for me. I will forever be thankful to Professor Randy Bradley. Randy has been tremendously helpful in completing my dissertation, and the job market. I am happy to acknowledge my debt to Professor Bogdan Bichescu, whose inspiring advise turned me towards writing my thesis in logic. I would like to thank Professor Neeraj Bharadwaj for his time and great advise.

I would also like to thank my college advisor, Professor Joonmo Cho. He was and remains my best role model for a researcher, teacher, and mentor.

## Abstract

In this dissertation, I present a new approach to capturing dependence across time in dynamic choice data. To achieve this, I develop a state space dynamic choice model and a novel algorithm to fit the data. Instead of capturing dependence in outcomes through lagged response variables, referred to as *state dependence,* I introduce a lagged utility term through the latent state equation. The lagged utility term captures *habit persistence,* which has not been explored directly in earlier models (Heckman, 1981b). The autoregressive nature of the lagged utility provides a significantly richer summary of prior utility than a lagged outcome variable. The fitting algorithm combines a non-linear particle filter with a standard Metropolis-Hastings step to compute Bayesian posterior estimates of the parameters. The model can capture habit persistence (inertia), variety seeking, serial correlation, and unobserved heterogeneity. Through simulation analysis, I demonstrate that while the proposed method is effective in estimating the parameters, both a large sample size and the number of simulated particles are critical. Misspecification in serial correlation in the random component of the utility function is shown to result in biased estimates for certain coefficients, although not the habit persistence term. This method avoids the initial conditions problem common with lagged variables (Wooldridge, 2010). From the perspective of a marketer, the value of the proposed model stems from its ability to distinguish the effects of habit, variety seeking, and heterogeneity.

The algorithm is applied to case studies involving the sales of fast-moving consumer goods, as recorded in scanner data furnished by a major grocery store. The studies demonstrate the wide-ranging variation in purchasing habits and price sensitivity across customers; this variation highlights the value of the individual-level models applied in this study. Specifically, I find the existence of habitual purchasing behavior in utilitarian goods (e.g., cereal and soft drinks). However, in hedonic goods (e.g., beer), I find no evidence of habit persistence, which is in agreement with earlier studies.

## Table of Contents

## List of Tables

## List of Figures

# Chapter 1

## Introduction

Over the past 30 years, numerous authors have attempted to extend the discrete choice model (Greene, 2003; McFadden, 1974) to analyze datasets where the researcher observes repeated choices from the same individuals or households over time. Perhaps the most obvious example of such data are repeated purchases of a product category in retail scanner data, but numerous other examples are available across a wide range of fields (Guadagni & Little, 1983; Heiss, 2008; Kitamura, 1990). These models are often referred to as dynamic choice models. The main challenge to modelling such data is capturing the correlation that exists over time in a non-linear regression model.

The goal of this dissertation is to investigate a new approach to modeling dependence across time in dynamic choice panel data. To accomplish this, I introduce a new state space approach to dynamic choice models and further provide a novel fitting method using sequential Monte Carlo methods. At its core, the method replaces the lagged outcome variables favored by previous models with a continuous latent state variable to control for the impact of previous experience. This represents a distinct break from most previous approaches to the dynamic discrete choice models. These models viewed the choice process in terms of a discrete state Markov chain via the use of lagged dependent variables to represent the prior state of the system (Keane, 2013; Seetharaman, 2004).

The introduction of this more flexible model comes at the cost of an increased computational burden but can now be overcome through the application of a modern

Monte Carlo simulation approach called the particle filter (Doucet, Godsill, & Andrieu, 2000). This approach has been commonly applied in nonlinear time series settings, particularly in financial and engineering applications. To my knowledge, this dissertation is the first work to apply the particle filter in the general choice model setting, allowing a variety of complex novel models to be fit once they have been translated into a state space formulation. This approach allows me to capture a variety of modeling effects, including correlated error terms and other latent effects, such as lagged utility, with relative ease.

As a concrete example of this methodology, I present two case studies that apply my approach to fast-moving consumer goods. These provide novel insights into individual consumer-level behavior, including price elasticity measures and customer inertia. I also point out the capability of the model to capture variety-seeking behavior in a natural way in contrast to many of the approaches considered in the literature. Finally, my work is the first to analyze inertia and variety-seeking behavior in comparing hedonic and utilitarian goods.

The above contributions are relevant and important for those social scientists and engineers who use dynamic choice models because they significantly enrich the capabilities of the models, provide improved accuracy, simplify the model implementation, and may offer insights that were not possible with earlier modelling strategies. In discussing choice models in the analysis of marketing and consumer behavior, my general interest is primarily applied statistical methodology, so I limit the focus of this dissertation to reduced form models.

In the remainder of this chapter, I provide a general introduction to dynamic choice models and a review of the literature in this area. This introduction and technical review builds a foundation for the technical results in subsequent chapters. Finally, I summarize the content of the remainder of the dissertation.

## 1.1 The Discrete Choice Model and Extensions

**1.1.1 Choice Models.** Discrete choice models have become an important tool for empirical studies since the pioneering work by Daniel McFadden in the 1970s and 1980s (Athey & Imbens, 2007; Hausman & McFadden, 1984). McFadden's early work focused on logit-based choice models in the area of transportation. Since that time, these models have been applied in many different domains of economics, such as labor economics, public finance, finance, marketing, as well as numerous other areas where the human decision-making process is involved (Athey & Imbens, 2007). Great advances in understanding dynamics of choice can be traced to the development of multinomial choice models (Ashok, Dillon, & Yuan, 2002; Erdem & Winer, 1998).

The impact of the tool to the study of economics was reinforced in 2000 when McFadden was awarded the Nobel Prize, which recognized in particular "his development of the theory and methods for analyzing discrete choice" (see Nobel 2000).

To understand both how and why the discrete choice model is useful, I consider a concrete example. McFadden (1974) used discrete choice models to describe how residents of Pittsburgh, Pennsylvania, chose a shopping destination. The outcome variable, *region*, was separated into five possible destinations based on city zones. One explanatory variable measures $S$ = shopping opportunities in the region based on the

number of retail jobs located there. A second variable, *P* = the price of the trip, is based on a separate study of the net costs of auto-in time and operating costs. For individual *i,* the odds of selecting each pair of choices (regions), *a* and *b* takes the logistic regression form

$$\log\left(\frac{Pr_a(s_{ia}, p_{ia})}{Pr_b(s_{ib}, p_{ib})}\right) = \beta_1(s_{1a} - s_{ib}) + \beta_2(p_{ia} - p_{ib}) \qquad (1.1)$$

where $\Pr_a(s_{ia}, p_{ia})$ represents the probability of an individual visiting a region with a particular cost and set of shopping opportunities. Although this appears to be a standard multinomial logit model (conditional logit), an important distinction is that the explanatory variables are actually attributes of the choice outcomes as opposed to being characteristics of the sampled individuals. In fact, both types of characteristics can be included in such models if the correct parameterization is used.

McFadden's derivation of this model utilized a latent variable approach. He assumed that each choice outcome provided a certain utility based on its features as well as the potential characteristics of the individual unit sampled. Considering the example above, let $y_{ij}^*$ represent the utility of shopping in region *j* for individual *i*. I could then represent the unobserved utility of this choice as

$$y_{ij}^* = \beta_1 S_{ij} + \beta_2 P_{ij} + \xi_{ij}\,, j = 1, \ldots, 5, \qquad (1.2)$$

where $\xi_{ij}$ represent unobservables affecting taste. The observed outcome is then assumed to be the choice which corresponds to the largest utility, $argmax_j\ (y_{i1}^*, \ldots, y_{iJ}^*)$. If I assume that $y_{ij}^*$ follows the type 1 extreme value distribution (EV1), then because differences in EV1 random variables follow the logistic distribution, the logistic regression model (1.1) results. In other words, the logit model is obtained by assuming that each $\xi_{ij}$ is independently, identically distributed with an extreme value distribution. The distribution is also called Gumbel and type I extreme value.

Based on the argument above, this model represented a major breakthrough in econometrics in that it allowed researchers to directly measure average differences in utility based on differences in choice characteristics. Both these theoretical characteristics, and the simple ability to estimate response probabilities as a function of choice characteristics, make this model of wide interest in both applied and theoretical settings as discussed above. Furthermore, standard logistic regression models can be viewed simply as a special case of these more general models.

Naturally, other distributions for the error terms can be substituted, and this is often done for convenience. In particular, replacing the EV1 distribution with the multivariate normal results in the multinomial probit (conditional probit) model, which has some advantages but is generally very similar to the conditional logit model presented here.

For further details, numerous textbook expositions are available, including Greene (2003), Wooldridge (2010), and Cameron and Trivedi (2005). Book-length treatments

are also available, including Train (2009) and Hensher, Rose, and Greene (2005), among others.

**1.1.2 Dynamic Choice Models.** Panel, longitudinal, or clustered data concerns the study of repeated observations of a sample from a population. In many settings, this data will appear as a collection of short- or medium-length time series. Analysis of such data can include analysis of between-series effects and within-series effects, which differentiate this data from time series data. If the dependent variable of interest in a panel data set is a discrete categorical or choice outcome, then the discrete choice model may be an appropriate tool.

A very common application of choice models to panel data occurs in the analysis of customer purchase dynamics in scanner panel data from grocery stores. Choice model analysis of this data first appeared in the pioneering paper of Guadagni & Little (1983). Numerous others have since used this data to analyze a variety of related topics; see Keane (2013) for a comprehensive review or P. K. Chintagunta (1999) for an alternative perspective.

To extend the discrete choice model to this context, I extend (1.1-1.2) using the identical utility equations. Let $i = 1, \dots, I$ denote an individual in a study, $j = 1, \dots, J$ represent a particular brand or product category choice, and $t = 1, \dots, T_i$ designate a particular time for the purchase choice. Here, neither the timing of the individual purchase *t* nor the number of observations $T_i$ are assumed to be common across individuals. Given this setup, I assume that the utility for a particular product at a given

time $y_{ijt}$ is the sum of a linear additive function of predictors $x_{ijtm},\ m = 1, \ldots, M$, and a latent term capturing unobserved taste factors $\xi_{ijt}$. Using these terms, I can then write the following utility function:

$$y^*_{ijt} = \beta_{j0} + \sum_{m=1}^{M} \beta_m \cdot x_{ijtm} + \xi_{ijt}. \tag{1.2}$$

Here, parameter $\beta_{j0}$ denotes the brand-specific constant for brand *j*, and $\beta_m$ is a vector of coefficients capturing the individual and choice attribute effects on the individual's evaluation of utility from brand $j$. Note that the explanatory variables $x_{ijtm}$ can represent both characteristics of the choice *j* as well as the individual *i.* In the latter case, the index *j* may be unnecessary, but I don't distinguish between these cases at this point.

As discussed earlier, the probability of the observed choice will depend upon the distribution of unobserved taste factors or shocks. If $\xi_{ijt}$ is i.i.d. extreme value, then a dynamic logistic regression model results in:

$$\Pr(y_{it} = j) = P_{ijt} = \frac{\exp\left(y^*_{ijt}\right)}{\sum_{j=1}^{J} exp\left(y^*_{ijt}\right)}, \tag{1.3}$$

while assuming normally distributed taste shocks produce a probit version of the model.

When considering a model of outcomes over time, i.e.. a panel data model, it is usually critical to consider the impact of previous outcomes for a unit (e.g., individual or household) on the current outcome. For example, in the context of repeat grocery product purchases, one of the most obvious features of the data is persistence of brand choice over time. Table 1.1 provides an example of persistence arising in the pancake

mix market using a subset of a database of scanner transactions occurring at noncompeting retail chains located across the United States and owned by a single firm. (This data is discussed further in Section 4.5.) It is clear from the diagonal entries that the probability of repurchasing the current brand on the next in-category purchase is much higher than 50% for almost all brands. Previous experience with a particular brand may weigh heavily on the current utility of various choices. If the goal of a particular study is to assess the impact of price or promotion on an individual's choice, then ignoring this prior behavior may produce biased estimates and incorrect inferences (Keane, 1997).

**Table 1.1 - Distribution of Pancake Mix**

| | **1.** | **2.** | **3.** | **4.** | **5.** | **6.** | **7.** | **8.** | **9.** | **10.** | **11.** | **12.** | **13.** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Frequency | 9,671 | 1,775 | 176 | 173 | 615 | 5,206 | 1,213 | 272 | 320 | 406 | 425 | 5,043 | 2,553 |
| Percent | 35% | 6% | .6% | .6% | 2% | 19% | 4% | 1% | 1% | 1% | 2% | 18% | 9% |
| **By previous transaction [%]** | | | | | | | | | | | | | |
| **1.** Aunt Jemima | **81.9** | 6.6 | | | 20 | 14.9 | 14.3 | | 16.7 | 20.8 | 3.8 | 9.2 | 3.6 |
| **2.** Bisquick | .5 | **83.2** | | | | 1.1 | | | 0.0 | 1.3 | | .4 | 2.2 |
| **3.** Bruce's | .2 | | **69.2** | | | | | | 4.2 | | | .4 | |
| **4.** Classique | .3 | 1.5 | | **87.5** | | .3 | | 5.6 | | | | .4 | |
| **5.** Hodgson Mill | .2 | | | | **60** | .9 | 1.6 | | 4.2 | | | .8 | |
| **6.** Hungry Jack | 7.4 | 4.4 | 7.7 | | 8 | **62.6** | 12.7 | | 4.2 | 14.3 | | 6.5 | 4.4 |
| **7.** Krusteaz | 1.0 | 2.2 | 7.7 | | | 2.6 | **46** | | 4.2 | 3.9 | 7.7 | 2.3 | .7 |
| **8.** MW Flap-Stax | .2 | | | | | | | **88.9** | | | 3.8 | .4 | |
| **9.** Maple Grove | | | 7.7 | | | | 1.6 | | **58.3** | | | | .7 |
| **10.** Mrs. Butterworth | 1.9 | | | | 4 | 6.0 | 3.2 | | 4.2 | **49.4** | | 3.1 | 2.2 |
| **11.** Pioneer | | | | | | | 1.6 | | 4.2 | | **73.1** | | .7 |
| **12.** Private Label | 5.2 | 1.5 | | 12.5 | 4 | 11.7 | 12.7 | 5.6 | | 10.4 | 11.5 | **75.9** | 1.5 |
| **13.** White Lily | 1.2 | .7 | 7.7 | | 4 | | 6.3 | | | | | .8 | **83.9** |

Guadagni and Little (1983) defined a smoothed historical purchase measure in order to control for previous purchase behavior in a logistic discrete choice model of scanner data. In their analysis, they noted heterogeneity across customers in brand choice.

They used the term "brand loyalty" to define their operational measure despite its inadequacy in capturing certain aspects of the careful definition of loyalty provided by Jacoby and Chestnut (1978). In that work, the "loyalty" $GL_{ijt}$ of consumer $i$ for brand $j$ prior to choice occasion $t$ of this consumer is determined by an exponentially smoothed weighted average of past purchases of the brand:

$$GL_{ijt} = \alpha \cdot GL_{ij,t-1} + (1-\alpha) \cdot y_{ij,t-1} \tag{1.3}$$

, where $0 \leq \alpha \leq 1$ and

$$y_{ijt} = \begin{cases} 1 \; if \; consumer \; i \; buys \; brand \; j \; in \; transaction \; t \\ 0 \qquad\qquad\qquad otherwise \end{cases}.$$

The constant $\alpha$ is an exponential smoothing parameter and indicates how loyalty grows for a chosen brand and declines for brands not chosen in the purchase occasion (Baumgartner, 2003).

Despite their modest goal of using this index to control for heterogeneity in the choice across shoppers, this measure became a de facto operational measure of both inertia and repurchase (correlation over time). For example, Corstjens and Lal (2000) note that "similarly, our notion of inertia is also captured by the loyalty measure proposed by Guadagni and Little (1983)." However, Kanetkar, Weinberg, and Weiss (1990) show that "the GL-index does not behave in a manner consistent with our common sense understanding of brand loyalty ... even though it plays an important predictive role in the

multinomial logit brand choice model" (McAlister et al., 1991). They also suggest that the GL term is not measuring brand loyalty. Instead, it is accounting for heterogeneity among households that stems from unobserved variables and manifests itself as first- and higher-order effects in the purchasing process. Hence, although the GL term may have been misappropriated as an operational definition of various terms, it seems to be a relevant tool for investigating persistence in repeat purchases, although its precise role requires clarification.

Contemporaneously with the work of Guadagni and Little, J.J. Heckman authored a series of groundbreaking articles investigating the use of dynamic models for choice (Heckman, 1977, 1981a, 1981b). These papers introduced and analyzed a general model for the role of observed past history in the analysis of choice. The principle interest was the unification of a large number of models applied to problems in labor economics. The general model proposes four sources of persistence in dynamic discrete probability models see Equation 3 of Heckman (1981a). Based on the structural choice of coefficients, this model can express a wide range of Markov and higher-order dynamical models.

Building on this earlier effort of Heckman and subsequent work by Gary Chamberlain (1984), a number of researchers adapted the model of Guadagni and Little (1983) to include the earlier theoretical contributions. Keane (2013) reviews the key model developments in this area, particularly as they relate to modelling of consumer demand. According to Keane, Heckman's work implies that there are three non-exclusive factors that can explain the observed heterogeneity in brand choice: 1)

permanent unobserved heterogeneity in tastes, 2) serial correlation in taste shocks, and/or 3) "true" or "structural" state dependence.

It is this third source of persistence, state dependence, that is of most consequence since it implies an effect of current choices on future choices (Keane, 2013). Uncovering whether state dependence exists is particularly important because, when using dynamic choice models to evaluate policy changes, for example, price discounts. The existence of state dependence will imply that current actions affect both current and future demand. It follows that if such an effect is not controlled for, it may bias any estimates of the effects of such policy changes.

Using the notation of Keane (2013), the typical structure of dynamic choice models lets $i = 1, \dots, I$ denote the unit, $t = 1, \dots, T$ index time, and $j = 1, \dots, J$ index the brand or product choice. Then, the model can be written as follows:

$$Y_{ijt}^{*} = \alpha_{ij} + x_{ijt} \cdot \beta + \phi \cdot d_{ij,t-1} + a_{ijt}, \qquad where\ a_{ijt} = \lambda \cdot a_{ij,t-1} + \xi_{ijt} \tag{1.4}$$

$$d_{ijt} = \begin{cases} 1 & if \quad Y_{ijt}^{*} > Y_{ikt}^{*} \ \ \forall k \neq j, \\ 0 & otherwise \end{cases} \tag{1.5}$$

Equation (1.4) expresses the utility that consumer $i$ receives from the purchase of brand $j$ at time $t$.

The first term in the utility expression depends on $\alpha_{ij}$, subject $i$'s intrinsic time invariant preference for brand $j$. The heterogeneity in the brand intercepts $\alpha_{ij}$ across categories capture a person's heterogeneity in tastes for attributes of choice that are not observed by data. Utility also depends on a vector of product attributes $X_{ijt}$ and the

attribute weights $\beta$; see Keane (2013) for further justification of the structure of the price effect, which would often be included in $x_{ijt}$. The term $a_{ijt}$ measures a "taste shock," and the model can consider idiosyncratic consumer, time and brand-specific taste. It is allowed to be serially correlated to capture the potential for time-varying tastes with the fundamental shock $\xi_{ijt}$ being independent and identically distributed and $\rho > 0$ indicating temporal persistence. As discussed in Keane (1997), it can be interpreted as either incomplete information on the part of the econometrician or as a result of unobserved brand attributes for which people have heterogeneous tastes that vary over time.

The term that most uniquely identifies this model as a dynamic model is the lagged outcome variable, $d_{ij,t-1}$. This indicates whether or not brand *j* was purchased at time *t-1*. The introduction of this term captures the effect of a lagged purchase of a brand on its current period utility. As mentioned, this effect is referred to as "structural" state dependence. This simple, lagged effect can be contrasted with the exponentially smoothed effect of the GL term described earlier. The consequence of including the term is that the model explicitly takes the form of a discrete state Markov chain – hence, the term state dependence. It is important to emphasize that the state described here is 1) discrete and 2) in the consumer demand setting refers to the last *observed* purchase without accounting for time since that purchase, not how that purchase would or should impact the next purchase. While in some settings, such as the labor market, it may be very clear how a prior state of employment may affect a subsequent choice, the effect may be less clear here. Keane (2013) notes a number of possible causes for discrete state

structural dependence of utility, such as habit persistence, learning, inventories, variety-seeking behavior, and switching costs.

**1.1.3 A Continuous State Space Dynamic Choice Model.** A major goal of this work is to introduce a new perspective on the dynamic choice model, which may be more appropriate conceptually and more flexible and interpretable than the previous discrete state space dynamic models in certain applications.

The concept of a state space model was introduced by celebrated electrical engineering professor Rudolf E. Kalman in a series of papers during the 1960s (Bucy, Kalman, & Selin, 1965; Rudolf E Kalman, 1962). The context in which the model was developed consisted of radar tracking and control of an object in time. To accomplish this, Kalman proposed the existence of a latent (unobserved) true state for the object, along with a historical series of data measurements and a current measurement. A simplified version of the situation can be represented simply with a system of equations:

$$y_t = Z_t \cdot a_t + \xi_t$$

$$a_{t+1} = T_t \cdot a_t + R_t \cdot \epsilon_t,$$

where $y_t$ is the tracked or measured position of the object; $Z_t$ is a mapping function that maps the state vector $a_t$ into the observed space; $T_t$ describes how the state changes over time; $R_t$ represents the control input variables, which are applied to the control vector to change the position of the object; and $\epsilon_t \; and \; \xi_t$ are random measurement errors. The

first equation describes the relation between the observed data and the "true" underlying state. The second equation describes how the state changes over time under both standard forces and control inputs.

While the description given here clearly focuses on an engineering application, over time it became clear to researchers in related fields, like statistics and economics, that the method offered considerable benefits in time-series modeling (Durbin & Koopman, 2012; Harvey, 1990; M West & Harrison, 1997).

Analogous to Keane's utility model, Equations 1.4 – 1.5, I propose the following state space utility model:

$$Y_{ijt}^{*} = a_{ijt} + \xi_{ijt} \tag{1.6}$$

$$a_{ijt} = \phi_i \cdot a_{ij,t-1} + \alpha_{ij} + x_{ijt} \cdot \beta + (\lambda_i + \phi_i) \cdot \epsilon_{ij,t-1}. \tag{1.7}$$

While the details of this model will be discussed in Chapter 2, this approach is distinct from the earlier model in several ways. First, the model proposed by Keane takes the form of a discrete state Markov process, with states being the *j* choice categories. As a consequence, it follows that the probability of making a particular choice depends upon the state that you are currently in, i.e., state dependence. In contrast, my latent state space model postulates the existence of a continuous underlying utility state for each category, which is modified when a purchase is made. It is apparent from Equation 1.6 that, given the current values of the latent states, $a_{ijt}$, the utility $Y_{ijt}^{*}$ is independent of previous values of $Y_{ij.}^{*}$, in particular $Y_{ijt-1}^{*}$. Given the value of the current continuous latent state, the current choice is independent of previous choices. Likewise, $\phi$ captures Heckman's

notion of habit persistence (Heckman, 1981b) and the essential idea in Coleman's "latent Markov" model that previous propensities to choose a state rather than previous occupancy of a state determine the current probability that a state is occupied (Coleman, 1964).

Several motivations exist for considering the continuous state space model described above. Foremost, the model may be conceptually more appropriate for certain problems, such as consumer demand. Although Keane (2013) asserts wide agreement that state dependence is a significant factor in habit persistence, he also notes that consumer taste heterogeneity is a much stronger source of observed persistence. By adding a latent term that models consumer taste continuously, this model may be able to better identify the carryover effects of earlier purchases on future patterns. In addition, even if only an approximation, such models offer a great deal of flexibility in specifying a variety of modelling structures, creating the opportunity to fit more realistic models (see Chapter 2 for further discussion).

Although Heckman (1981b) introduced a lagged utility term in his proposed dynamic choice model that was intended to capture "habit persistence," few subsequent researchers in demand modelling did not attempt to include this term, possibly due to computational difficulties. The idea of a continuous state space dynamic choice model is quite novel, having been considered earlier only by Heiss (2008) and in a somewhat modified form by Seetharaman (2004). Heiss (2008) avoided conceptual development and theoretical considerations of model structure and instead focused on numerical and simulation algorithms for fitting the model. Applications focused on repeated binary

self-assessment of personal health as opposed to consumer demand. Seetharaman (2004) mixes together the concept of latent utility with a number of other complex assumptions. A single additional article by McCormick, Raftery, Madigan, and Burd (2012) created a simple, dynamic, logistic regression model for binary data with the focus on model averaging for classification (prediction), as opposed to interpretation, which is the task considered here.

An aspect of these models that has not been touched on yet is correlation in error terms. When including lagged dependent variables in a model with correlated error terms, the error term and lagged dependent variable are correlated, violating the typical assumptions of the model and leading to inconsistency. Simulation experiments from Hsiao, Hashem Pesaran, and Kamil Tahmiscioglu (2002) strongly support this conclusion. To achieve consistent estimates in dynamic panel data, I need to correctly specify the serial correlation (more details in Section 3.2). In previous literature in this area, many authors have not controlled for this factor. Among the conventional panel data models, only Kean's (1997, 2013) model considered serial correlation in utilities. One goal of his study was to give a taxonomy of types of heterogeneity. Keane (1997) analyzed seven different types of models: (i) observed and unobserved heterogeneity in tastes for observed attributes, (ii) observed heterogeneity in brand intercepts, (iii) unobserved heterogeneity in tastes for unobserved common and unique attributes for which consumers have fixed tastes, and (iv) the same for attributes where consumers have time-varying tastes. He found that most of the observed persistence in alternative choice does appear to be due to taste heterogeneity, but there is still a significant effect

from state dependence. It is important to note that serial correlation in utility can be caused by serial correlation in unobserved components (i.e., error terms). However, error terms can be serially correlated due to serially correlated exogenous shocks to utility, which is distinct from the habit persistence introduced by Heckman (1981b). Because of this, Seetharaman's (2004) type 1 – habit persistence is simply a measure of serial correlation of error in utilities (rather than a measure of habit persistence).

In Table 1.2, I present a detailed comparison of previously proposed dynamic models in terms of the factors reviewed above: 1) incorporation of structural state dependence terms, 2) incorporation of serial correlation in utilities, and 3) incorporation of habit persistence through lagged utility. The table discusses *lightning bolt* (LB) models, which were introduced by Roy, Chintagunta, and Haldar (1996) and further utilized by Chintagunta (P. K. Chintagunta, 1998, 1999, 2001) and Seetharaman (2004). These models are consistent with the theory of random utility maximization of consumer choice behavior, and the underlying random utility process is Markov. The inter-temporal evolution of the brand choice process is also Markov. The most important distinction between the conventional dynamic panel models and the lightning bolt (LB) models is that LB models assume the consumer has limited recall capabilities and her current preference evaluation solely through the greatest of the unobserved signals. Chapter 2 discusses these models in greater detail.

**Table 1.2 - Models of State Dependence in Literature**

| Study | Structural State Dependence | Serial Correlation | Habit Persistence | Inconsistent Estimates (Spurious State Dependence or Habit Persistence) |
|---|---|---|---|---|
| Conventional panel data models | | | | |
| Guadagni and Little (1983) | Yes | No | No | Yes |
| Erdem (1996) | Yes | No | No | Yes |
| Gupta et al. (1997) | Yes | No | No | Yes |
| Keane (1997, 2013) | **Yes** | **Yes** | No | **No** |
| Seetharaman et al. (1999) | Yes | No | No | Yes |
| Abramson et al. (2000) | Yes | No | No | Yes |
| Erdem and Sun (2001) | Yes | No | No | Yes |
| Dube et al. (2008) | Yes | No | No | Yes |
| "Lightning Bolt" type models | | | | |
| Roy et al. (1996) | Yes | Yes | Yes | No |
| Seetharaman (2004) | Yes | Yes | Yes | No |

This literature review has focused on specific aspects of dynamic choice models that are most relevant to the contributions of this dissertation. Naturally, in such a mature field, numerous other factors have been identified as playing a key role in decision-making processes, and these factors will depend heavily on the field of study. For instance, in the marketing literature, factors such as price, placement, packaging, advertising, and choice set limitations and variability all have significant impact on the choice process (Andrews, Ainslie, & Currim, 2008; P. Chintagunta, Dubé, & Goh, 2005; Erdem & Keane, 1996; Keane, 2013). Many of these factors, if known and recorded, can be included in the models discussed here, while others may require more advanced structural modelling techniques. Structural models are reviewed in P. Chintagunta et al. (2005) and Chandukala, Kim, Otter, Rossi, and Allenby (2007) and provide a comprehensive review of choice models in marketing.

**1.2 Summary of the Dissertation**

Chapter 2 begins by introducing the general concept of the state space model and relates it to the random intercept model in panel data. I then present the state space version of the dynamic choice model and develop the likelihood function. The particle-filtering approach for sequential importance sampling is then introduced in order to implement and simulate the likelihood of Bayesian Monte Carlo inference. After the algorithm development, I discuss several of the virtues of this model, such as lack of dependence on initial conditions, ease of including numerous modelling effects, and issues with consistency of models under misspecification when models do and do not contain lagged dependent variables.

Chapter 3 explores three simulations testing the proposed technique using Monte Carlo simulation under a variety of data set and simulation replication sample sizes. The first simulation experiment is performed to demonstrate that the algorithm developed can accurately estimate a model containing only an intercept and a habit persistence, lagged utility, term. The second simulation tests the ability of the algorithm to accurately estimate both habit persistence and price sensitivity. The final experiment tests habit persistence, price sensitivity, and the effect of serial correlation on model fitting.

Chapter 4 reviews the concepts of state dependence and habit persistence and compares these factors in the analysis of supermarket scanner data. To begin, the concept of habit persistence is considered, and the operationalization of this concept through a lagged utility function is proposed in agreement with the earlier work of Heckman (1981b). State dependence is then discussed as a form of feedback and again is

operationalized through the use of lagged dependent variables, including the GL term mentioned above. These two concepts are explored and contrasted in terms of the information that they capture when modelling. Suitability to the marketing context is also deliberated. Finally, two brief case studies are presented which demonstrate both the flexibility and advantages of using the habit persistence term in comparison with the state dependence approach. The case studies also highlight important differences in consumer behavior within product categories and across products that are described as hedonic and utilitarian.

To conclude, Chapter 5 provides an extensive review of the key contributions of the dissertation along with a discussion of the limitations. Further development of this methodology is considered along with applications in marketing and management.

# Chapter 2

# State Space Models and Fitting Methods

The primary task of panel data analysis is to uncover the dynamics behind the evolution of observations measured over time for a population of individuals. In this chapter, I introduce the state space model for time series analysis and use it to construct a flexible model to capture discrete choice outcomes over time. The model that I develop is able to capture the effects of both observable and unobservable factors and model effects at both the individual and panel level.

To fit this model to actual dynamic data, I consider a new simulation method, called the particle filter, which uses sequential Monte Carlo importance sampling to fit the model to dynamic data (Doucet, De Freitas, & Gordon, 2001; Lopes & Tsay, 2011) and evaluate the quality of the fit. I contrast the proposed method with more traditional dynamic choice models with lagged variables and discuss the issues of model misspecification and the difficulty of controlling for initial conditions.

## 2.1 State Space Models

State space models are used extensively to model both financial time series and econometrics data (Commandeur & Koopman, 2007; Durbin & Koopman, 2012; Harvey, 1990; Tsay, 2005); nonetheless, they have a significantly longer history in engineering, where they are used to track and control the evolution of a given system, such as a missile (Kailath, Sayed, & Hassibi, 2000; Ra E Kalman, 1962).

For pedagogical reasons, I begin by discussing the linear Gaussian state space model, which was introduced in Section 1.1.3. Traditionally, this model is used to

analyze repeated observations, over a given period of time, which are assumed to depend linearly on an underlying latent state that is generated by a dynamic process (stochastically time-varying system). I assume that the observations are subject to measurement error and this error is independent of the state process. The linear state space model can be represented in several ways. In line with (Commandeur & Koopman, 2007; Commandeur, Koopman, & Ooms, 2011), I represent the state and observation equations respectively as

$$\boldsymbol{a}_{t+1} = \boldsymbol{T}_t \cdot \boldsymbol{a}_t + \boldsymbol{R}_t \cdot \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim N(0, \boldsymbol{Q}_t) \tag{2.1.1}$$

$$\boldsymbol{y}_t = \boldsymbol{z}_t \cdot \boldsymbol{a}_t + \boldsymbol{\xi}_t, \quad \boldsymbol{\xi}_t \sim N(0, \boldsymbol{H}_t), \tag{2.1.2}$$

where $\boldsymbol{a}_t$ is the state transition vector, $\boldsymbol{\xi}_{it}$ and $\boldsymbol{\epsilon}_t$ are disturbance vectors, and the system matrices, $\boldsymbol{z}_t, \boldsymbol{T}_t, \boldsymbol{R}_t$, and $\boldsymbol{Q}_t$, are fixed and known but a selection of elements may depend on unknown parameters. I use bold lowercase (uppercase) letters to denote a vector (matrix). Equation 2.1.1 is called the state transition equation, while Equation 2.1.2 is called the observation or measurement equation. The $p \times 1$ observation sequence, $\boldsymbol{y}_t$, contains the $p$ observations at time $t$, while the $m \times 1$ state transition, $\boldsymbol{a}_t$, is latent. The $p \times 1$ irregular vector, $\boldsymbol{\xi}_t$, has zero mean and the $p \times p$ variance matrix $\boldsymbol{H}_t$. The $p \times m$ matrix, $\boldsymbol{Z}_t$, links the observation sequence, $\boldsymbol{y}_t$, with the unobservable state transition, $\boldsymbol{a}_t$, and may consist of regression variables. The $r \times 1$ disturbance vector, $\boldsymbol{\epsilon}_{it}$, for the state transition, has a zero mean and the $r \times r$ variance matrix $\boldsymbol{Q}_t$. The $m \times m$ transition matrix, $\boldsymbol{T}_t$, determines the dynamic process of the state transition. I assume that the observation and latent state disturbances, $\boldsymbol{\epsilon}_t$ and $\boldsymbol{\xi}_t$, are serially independent and

independent of each other at all time points. Furthermore, the initial state transition $a_1$ is assumed to be generated as

$$a_1 \sim N(a_1', P_1), \tag{2.1.3}$$

and is independent of $\boldsymbol{\epsilon}_t$ and $\boldsymbol{\xi}_t$. The mean $a_1'$ and variance $P_1$ can be treated as given and known in most situations.

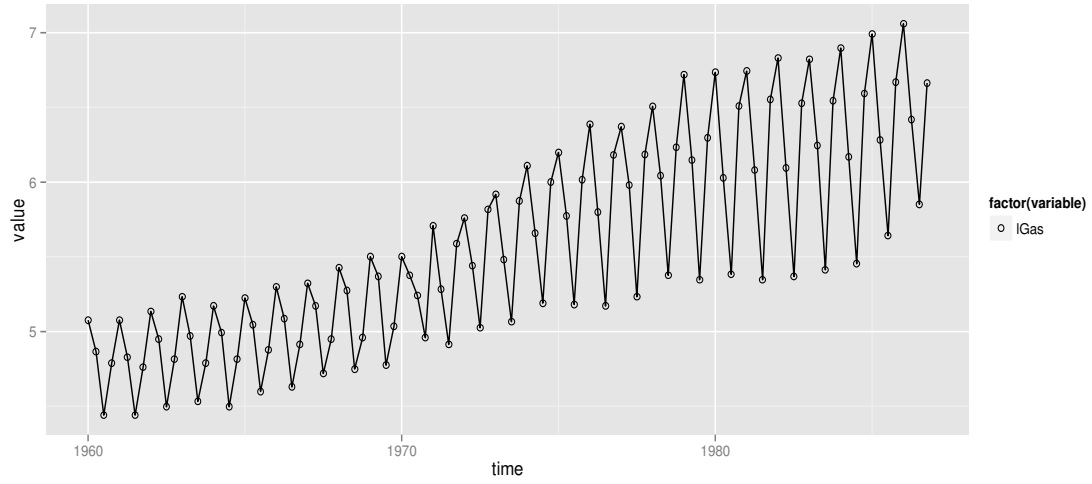



**Figure 2.1 - UK Gas Consumption**

Figure 2.1 shows the time series plot of UK gas consumption from 1960 to 1986 as an example (Durbin & Koopman, 2012). Suppose I want to describe gas consumption over time by a combination of a linear trend and a quarterly seasonal component. This is easily accomplished within the linear state space model framework described above.

The following equations describe the process: where $y_t$ represents the UK gas consumption, $\mu_t$ is the underlying level in prices, $v_t$ is a trend component, $\gamma_{1,t}, \gamma_{2,t}$,

and $\gamma_{3,t}$ represent the seasonal components, and $\xi_t$ and $\zeta_t$ represent the independent scalar error components, as described earlier:

$$\begin{aligned} y_t &= \mu_t + \gamma_t + \xi_t, \quad \xi_t \sim N(0, \sigma_\xi^2) \\ \mu_{t+1} &= \mu_t + v_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma_\varepsilon^2) \\ v_{t+1} &= v_t + \zeta_t, \quad \zeta_t \sim N(0, \sigma_\zeta^2) \\ \gamma_{1,t+1} &= -\gamma_{1,t} - \gamma_{2,t} - \gamma_{3,t} + \omega_t, \quad \omega_t \sim N(0, \sigma_\omega^2) \\ \gamma_{2,t+1} &= \gamma_{1,t}, \\ \gamma_{3,t+1} &= \gamma_{2,t}. \end{aligned} \tag{2.1.4}$$

Equation 2.1.4 can be expressed in terms of equations 2.1.1 and 2.12 using the following matrix terms:

$$\boldsymbol{a}_t = \begin{pmatrix} \mu_t \\ v_t \\ \gamma_{1,t} \\ \gamma_{2,t} \\ \gamma_{3,t} \end{pmatrix}, \quad \boldsymbol{\epsilon}_t = \begin{pmatrix} \epsilon_t \\ \zeta_t \\ \omega_t \end{pmatrix}, \quad \boldsymbol{T}_t = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & -1 & -1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}, \tag{2.1.5}$$

$$\boldsymbol{z}_t = (1 \quad 1 \quad 0 \quad 0 \quad 0),$$

$$\boldsymbol{H}_t = \sigma_\epsilon^2, \quad \boldsymbol{Q}_t = \begin{bmatrix} \sigma_\varepsilon^2 & 0 & 0 \\ 0 & \sigma_\zeta^2 & 0 \\ 0 & 0 & \sigma_w^2 \end{bmatrix}, \text{ and } \boldsymbol{R}_t = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

In this model, the local linear trend model requires a $2 \times 1$ state transition, $\boldsymbol{a}_t$, one element for the level component, $\mu_t$, and the other element for the slope component, $v_t$.

This slope component is viewed as a time varying version of the regression coefficient in the classical linear trend model: $y_t = \mu + v \cdot t + \xi_t$ with $\mu = \mu_1$ and $v = v_1$. Based on the fitted model, the smoothed estimates illustrate a decomposition of the observed data into a smooth trend and a seasonal stochastic component. In the estimation, I assumed that the trend follows an integrated random walk process (i.e., $\sigma_\xi^2 = 0$). Figure 2.2 shows smoothed estimates of trend and the stochastic seasonal component.

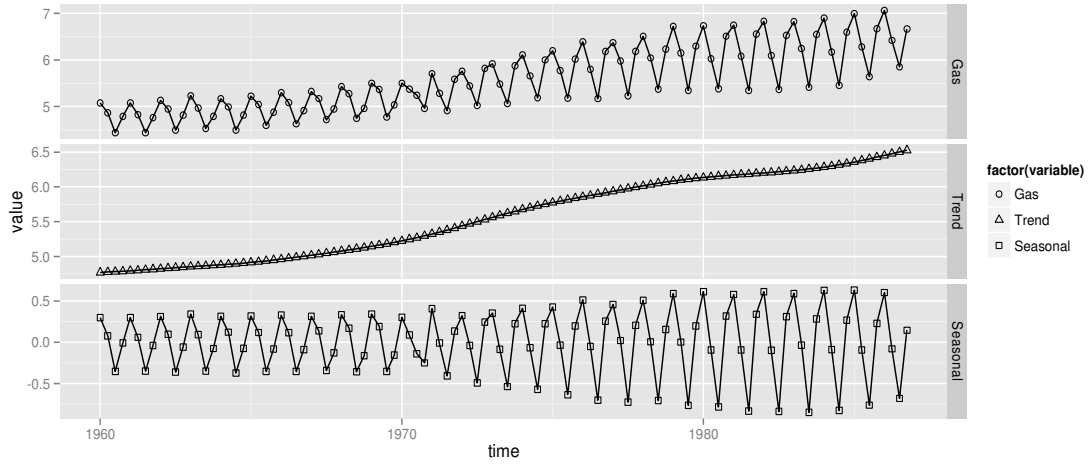



**Figure 2.2 - Estimates of Trend and Stochastic Seasonal Component**

While state space models are by no means the only approach for modelling data that evolves in time, Durbin and Koopman (2012) point out a number of advantages of these models. First, the key advantage of the approach is that it is based on a structural analysis of the problem. Second, the different components that make up the series, such as trend, seasonal, cycle, and calendar variations, combined with the effects of the explanatory variables and interventions, are modelled separately before being put

together in the state space model. In this model, as in a regression model, the investigator identifies and models any features. (In contrast, the better Box-Jenkins approach is a kind of “black box” in which the adopted model depends purely on the data without prior analysis of the structure of the system. In particular, I gain no knowledge of the factors that drive the system, only an ability to forecast.) Third, the flexibility described above makes it easy to accommodate known changes to the system over time. As an example, in my case, I may wish to allow for promotions and price changes at known times. Finally, state space models allow easier treatment of missing observations, explanatory variables can be incorporated, regression coefficients can be allowed to vary over time, calendar variations such as store closings for holidays can easily be addressed, and no extra theory is required for forecasting in general, since it follows easily from Bayes theorem. Furthermore, the above points apply equally well to linear Gaussian models and nonlinear models, such as the discrete choice model, which is the focus of my interest.

While state space models have been considered extensively in the context of Gaussian models, fewer attempts have been made to adopt them in a nonlinear univariate series and fewer still in nonlinear panel models. Among them, Durbin and Koopman (2000) survey non-linear state space models for a univariate time series. They consider non-Gaussian models in both the state and the observation equations. They use a linearization method to approach the problem and discuss simulation methods based on importance sampling, which are related to the methods introduced later in this section. In an earlier effort, Mike West, Harrison, and Migon (1985) introduced dynamic generalized linear models along with a Bayesian fitting method.

More recently, McCormick et al. (2012) studied dynamic logistic regression and inference based on model averaging. They proposed an online binary classification procedure for cases when there is uncertainty about the proper model to use and when the parameters within a model change over time. Uncertainty was accounted for through a dynamic extension of Bayesian Model Averaging (BMA) in which posterior model probabilities are also allowed to change with time. The goal of the McCormick et al. study was to determine the accurate classification of medical outcomes. This is distinct from my own research interests, which focus on model interpretation.

Frühwirth-Schnatter and Frühwirth (2007) propose a mixture sampling scheme that is useful in fitting a particular type of state space model, dynamic logistic regression with time varying regression coefficients. Their focus is on a Monte Carlo sampling algorithm, which allows efficient Bayesian inference for a small subset of state space models.

In the context of nonlinear panel data, Heiss (2008) has explored a state space approach for capturing dynamic model behavior in numerous small time series. In doing so, he developed non-linear filtering algorithms appropriate for non-linear panel data models with autoregressive error components.

## 2.1 A State Space Discrete Choice Model

### 2.2.1 Deriving the Choice Model Structure

Derivations for the multinomial probability structure of the discrete choice model are widely available in econometrics texts such as in the work of Train (2009), Greene

(2013), and Wooldridge (2010). In the approach that follows, I use the logistic form of the model adopted by Allenby and Lenk (1994). For continuity and a point of reference, I provide a derivation of the model. Readers with knowledge of this model can move directly to Section 2.3, where the state space version of the choice model is introduced.

Return to the dynamic discrete choice model introduced in Section 1.1.2, where $y_{it} = j$ denotes the event that individual $i$ chooses choice $j$ at time $t$. Then, individual $i$'s utility for choice $j$ at time $t$ is

$$Y_{ijt}^{*} = a_{ijt} + \xi_{ijt}, \tag{2.2.1}$$

where $a_{ijt}$ is a subject, choice, and time specific intercept, and $\xi_{ijt}$ captures unobserved choice specific features for the individual $i$ and choice $j$ at time $t$. The researcher does not observe $\xi_{ijt}$ for all of $j$; therefore, I treat these terms as random. I denote $f(\xi_{ijt})$ as the joint density of the random vector $\boldsymbol{\xi}_{\boldsymbol{i,t}}' = (\xi_{i1t}, \ldots, \xi_{iJt})$. Dropping the fixed factor for now, I can calculate the probability that the individual $i$ makes the choice $j$ with the density:

$$\begin{aligned} p_{ijt} &= p(U_{ijt} > U_{ikt}, \ \forall j \neq k) \\ &= p(a_{ijt} + \xi_{ijt} > a_{ikt} + \xi_{ikt}, \ \forall j \neq k) \\ &= p(\xi_{ijt} - \xi_{ikt} < a_{ikt} - a_{ijt}, \ \forall j \neq k). \end{aligned} \tag{2.2.2}$$

By using the density $f(\xi_{i.t})$, the cumulative probability of Equation 2.2.2 can be rewritten as

$$p_{ijt} = p\left(\xi_{ijt} - \xi_{ikt} < a_{ikt} - a_{ijt},\ \ \forall j \neq k\right) \tag{2.2.3}$$

$$= \int I\left(\xi_{ijt} - \xi_{ikt} < a_{ikt} - a_{ijt},\ \ \forall j \neq k\right) \cdot f(\xi_{i.t}) d\xi_{i.t},$$

where $I(\cdot)$ is the indicator function, equaling 1 if the expression in the parentheses is true and 0 otherwise. This is a multidimensional integral over the product of densities $f(\xi_{i.t})$. If the $\xi_{ijt}$ are independent and identically distributed with a type I extreme value distribution, the logit model is obtained, since the difference, $\xi^*_{ijkt} = \xi_{ijt} - \xi_{ikt}$, follows a logistic distribution,

$$F\left(\xi^*_{ijkt}\right) = \frac{\xi^*_{ijkt}}{1 + \xi^*_{ijkt}}. \tag{2.2.4}$$

The extreme value distribution has slightly fatter tails than normal and one might expect that this would lead to slightly more irrational behavior in purchasing than the normal distribution. However, Train (2009) notes, “the difference between extreme value and independent normal errors is indistinguishable empirically.”

Following McFadden (1974), I now can derive the logit choice probabilities. According to Equation 2.2.2, I have the probability that the individual $i$ makes the choice $j$ at time $t$ as

$$P_{ijt} = P\left(a_{ijt} + \xi_{ijt} > a_{ikt} + \xi_{ikt},\ \ \forall j \neq k\right).$$
$$= P\left(\xi_{ikt} < a_{ijt} - a_{ikt} + \xi_{ijt},\ \ \forall j \neq k\right).$$

Because each unobserved component of utility follows a type I extreme value density, the expression above becomes

$$= e^{-e^{-a_{ijt}+a_{ikt}-\xi_{ijt}}}. \tag{2.2.5}$$

Since I assume that $\xi_{ijt}$ is considered as given and the $\xi$'s are independent, the cumulative distribution over $\forall j \neq k$ is the product of the individual cumulative distributions:

$$P_{ijt} = \prod_{j \neq k} e^{-e^{-a_{ijt}+a_{ikt}-\xi_{ijt}}}. \tag{2.2.6}$$

However, $\xi_{ijt}$ is not given in a practical situation, I integrate out the unobserved portion of utility, $\xi_{ijt}$. Therefore, the choice probability is the integral of $p_{ijt}$, given $\xi_{ijt}$, over all values of $\xi_{ijt}$, weighted by its density, Equation 2.1.11:

$$P_{ijt} = \int \left( \prod_{j \neq k} e^{-e^{-a_{ijt}+a_{ikt}-\xi_{ijt}}} \right) e^{-\xi_{ijt}} e^{-e^{-\xi_{ijt}}} d\xi_{ijt}. \tag{2.2.7}$$

This leads to a succinct, closed form expression by applying some algebraic manipulation of the integral in Equation 2.1.16:

$$P_{ijt} = \frac{\exp(a_{ijt})}{\sum_{j=1}^{J} \exp(a_{ijt})}. \tag{2.2.8}$$

Allenby and Lenk (1994) adopted this functional form for utility and expanded $a_{ijt}$ to include both fixed effect terms, which are choice dependent, as well as random unobserved terms, which follow a normal distribution. Although this seems redundant given the model formulation was developed based on assuming type-1 extreme value errors, it will be convenient from a modeling perspective to allow a second set of time and choice specific random errors of normal form, which can be used to capture autocorrelation in unobserved aspects of the utility.

### 2.2.2 Including a State Space Component in the Dynamic Discrete Choice Model

To motivate the development of the state space dynamic discrete choice model, first consider the much simpler discrete choice model with a random effect. Consider a general discrete choice model with a random intercept

$$y_{ijt}^{*} = a_i + \beta_i \cdot x_{ijt} + \xi_{ijt} \tag{2.2.9}$$

$$p_{ijt} = \frac{\exp\left(a_i + \beta_i \cdot x_{ijt}\right)}{\sum_k \exp\left(a_i + \beta_i \cdot x_{ikt}\right)}, \tag{2.2.10}$$

where $i = 1, \ldots, N$ indicates individuals, $t = 1, \ldots, T$ indicates the time of a measurement, $y_{ijt}^{\star}$ is the utility of the choice, and $p_{ijt}$ indicates the probability of making the choice; see Wooldridge (2010, Section 16.2.2) . The unobserved variable $y_{ijt}^{*}$ is a function of a vector of covariates, $x_{ijt}$, which may contain time-varying, strictly exogenous variables, a time constant individual random intercept, $a_i$, and an i.i.d. error term $\xi_{ijt}$. I assume that random intercept and error terms are mutually independent, independent of $x_{ijt}$, and have a known parametric distribution. Hence, the observed outcome (choice) variable, $y_{ijt}$, is a parametric function of these unobserved variables.

In this study, I consider the generalizations of this class of models (i.e., state space models). From Equation 2.1.10, I replace the time constant component, $a_i$, with a laternt first order Markov process, $a_{ijt} = \phi_i \cdot a_{ij,t-1} + w_{ijt}$. Therefore, a random intercept model is a special case of state space model with $\phi_i = 1$ and no error term.

To extend the model from a static individual specific random effect to a general time varying state space approach, assume that $a_{ijt}$ can be decomposed into four parts:

$$a_{ijt} = \phi_i \cdot a_{ij,t-1} + \alpha_{ij} + \beta_i \cdot x_{jt} + w_{ijt}$$
$$w_{ijt} = \lambda_i \cdot w_{ij,t-1} + \eta_{ijt}, \qquad \eta_{ijt} \sim N(0, \sigma_\eta), \tag{2.2.11}$$

where $\beta_i$ is again a fixed utility weight of the product attributes $x_{ijt}$ for consumer $i$ at time $t$, $\alpha_{ij}$ indicates consumer $i$'s intrinsic preference for brand $j$, $\phi_i$ dictates how the utility at the current state depends on the previous latent state $a_{ij,t-1}$, and $w_{ijt}$ is a serially correlated error term with the fundamental shock $\eta_{ijt}$ being $iid$. The outcome, or choice Equation 2.2.11, simply says that individual $i$ chooses brand $j$, which gives her the maximum utility at purchase occasion time $t$. For simplicity, I write the probability in Equation 2.2.10 as a function of these terms:

$$\begin{aligned} p_{ijt} &= p\left(y_{ijt} = j \middle| \phi_i, \beta_i, \lambda_i, \alpha_{ij}, x_{ijt}\right) \\ &= p\left(y_{ijt} = j \middle| \theta_i, x_{ijt}\right), \end{aligned} \tag{2.2.12}$$

where $\theta_i = (\phi_i, \beta_i, \lambda_i, \alpha_{i.})$.

Furthermore, if I define the above state space choice model 2.2.1 as

$$Y_{ijt}^* = a_{ijt} + w_{ijt} \tag{2.2.13}$$
$$a_{ijt} = \phi_i \cdot a_{ij,t-1} + \alpha_{ij} + \beta_i \cdot x_{ijt} + (\lambda_i + \phi_i) \cdot w_{ij,t-1} + \xi_{ijt},$$

this model is equivalent to a dynamic regression with $ARMA(1,1)$ errors (an ARMA(p,q) model has the form $y_t = \sum_{j=1}^{p} \phi_j y_{t-j} + w_t + \sum_{j=1}^{q} \lambda_j w_{t-j}$). To validate this, I substitute the second equation into the first and derive the recurrence relation:

$$
\begin{aligned}
Y_{ijt}^{*} &= a_{ijt} + w_{ijt} \\
&= \phi_i \cdot a_{ij,t-1} + \alpha_{ij} + \beta_i \cdot x_{ijt} + (\lambda_i + \phi_i) \cdot w_{ij,t-1} + w_{ijt} + \xi_{ijt} \\
&= \alpha_{ij} + \beta_i \cdot x_{ijt} + \phi_i \cdot (a_{ij,t-1} + w_{ij,t-1}) + \lambda_i \cdot w_{ij,t-1} + w_{ijt} + \xi_{ijt} \\
&= \alpha_{ij} + \beta_i \cdot x_{ijt} + \phi_i \cdot Y_{ij,t-1}^{*} + \lambda_i \cdot w_{ij,t-1} + w_{ijt} + \xi_{ijt},
\end{aligned}
$$

demonstrating that the model is equivalent to an autoregressive form on the latent utility variable (Akaike, 1974).

### 2.3 Likelihood and Model Inference

This section introduces the challenges and specifics of fitting the proposed nonlinear state space model to data. I begin by deriving the form of the likelihood function theoretically. I then discuss Bayesian inference techniques for estimating the key parameters before exploring the particle filter and other approaches to non-linear filtering and numerical integration.

#### 2.3.1 Computing the Likelihood, Non-linear Filtering of the State Space Model

Statistical inference, whether Bayesian or frequentist, requires us to identify the likelihood function. Because the model outlined in Equation 2.2.11 continues unobserved

latent terms, the state, $a_{ijt}$, depends upon the error terms $w_{ijt}$, I must develop an approach to integrate over these values to compute the likelihood. In the following, I suppressed the individual subscripts, $i$, to simplify notation. To develop the likelihood, first assume that $y_{1:T} = (y_1, \ldots, y_T)$, $a_{1:T} = (a_1, \ldots, a_T)$, and $w_{0:T-1} = (w_0, \ldots, w_{T-1})$ are all observed. Then, the complete data joint distribution can be written as

$$
\begin{aligned}
&Q(y_{1:T}, a_{1:T}, w_{0:T-1} |\, x;\, \theta) \\
&\quad = P(y_{1:T}|a_{1:T}, w_{0:T-1}, x;\, \theta) P(a_{1:T}|w_{0:T-1}, x;\, \theta) P(w_{0:T-1}|x;\, \theta) \\
&\quad = P(y_{1:T}|a_{1:T}, w_{0:T-1}, x;\, \theta) P(a_{1:T}|w_{0:T-1}, x;\, \theta) P(w_{0:T-1}; \theta), \qquad (2.3.1)
\end{aligned}
$$

with the second equality following from the fact that the vector, $w_{0:T-1}$, is independent of $x$.

Because $P(a_{1:T}|w_{0:T-1}, x;\, \theta) = P(a_0) \prod_{t=1}^{T} P(a_t|a_{t-1}, w_t, x, ;\, \theta)$, the above expression further reduces to

$$
= P(y_{1:T}|a_{1:T}, w_{0:T-1}, x;\, \theta) P(a_0) \prod_{t=1}^{T} P(a_t|a_{t-1}, w_{t-1}, x; \theta)\, P(w_{0:T-1}; \theta). \qquad (2.3.2)
$$

Next, I replace the third term, using the identity

$P(w_{0:T-1}; \theta) = P(w_0) \prod_{t=2}^{T} P(w_{t-1}|w_{t-2};\, \theta)$, resulting in

$$P(y_{1:T}|a_{1:T}, w_{0:T-1}, x, ;\ \theta)P(a_0)\prod_{t=1}^{T}P(a_t|a_{t-1}, w_{t-1}, x;\ \theta)P(w_0)\prod_{t=2}^{T}P(w_{t-1}|w_{t-2};\ \theta). \quad (2.3.3)$$

Finally, representing Equation 2.3.3 more compactly, I refer to Equation 2.3.4 as the complete data joint distribution:

$$Q(y_{1:T}, a_{1:T}, w_{0:T-1}|x;\ \theta)$$

$$= P(a_0)\prod_{t=1}^{T}P(y_t|a_t, w_{t-1}, x; \theta)P(a_t|a_{t-1}, w_{t-1}, x; \theta)P(w_{t-1}|w_{t-2}; \theta), \quad (2.3.4)$$

where $P(w_{t-1}|w_{t-2}; \theta) = P(w_0)$ when $t = 1$.

To derive the likelihood function, a function of only data and fixed parameters, $\theta$, this expression must be integrated over the pairs of unobserved random elements $(a_t, w_t), t = 0, \dots, T$

$$L_i(\theta) = \prod_{t=1}^{T}P(\ y_t|x; \theta) = \iint Q(y_{1:T}, a_{1:T}, w_{0:T-1}|x;\ \theta)da_{1:T}dw_{0:T-1}. \quad (2.3.5)$$

See Lopes and Tsay (2011), Durbin and Koopman (2000, 2012), Carter and Kohn (1999) (Carter & Kohn, 1994); Gerlach, Carter, and Kohn (2000), and Mike West (1987) for further discussions of filtering methods and likelihood computation in nonlinear state space models.

Calculating the likelihood function above for the choice model cannot be done in closed form and it requires the use of numerical quadrature or Monte Carlo integration techniques. In either case, the computational burden is intense and the development of algorithms written in a compiled language, along with multiprocessor implementations, may be necessary for the implementation to be of practical use. Heiss and Winschel (Heiss, 2008; Heiss & Winschel, 2008) discusses several strategies for computing such likelihoods in both time series and panel data contexts; I discuss these methods further in Section 2.3.2 below. Before exploring integration techniques for the likelihood more deeply, I discuss inference for the key model parameters.

### 2.3.2 Bayesian Inference

As with any statistical model, a number of forms of inference and prediction may be applied. Whether one is interested in filtering, forecasting, estimating marginal effects, or testing hypotheses, a key step is computing the estimates of the parameter vector $\theta$, which, in this case, contain information on covariate effects, habit persistence, and correlation in unobserved effects. If the likelihood function can be evaluated (approximately) using either simulation or alternative numerical methods, the Bayesian inference for $\theta$ can be implemented through traditional Markov Chain Monte Carlo techniques.

In the context of my study, Gibbs sampling is not possible because of the lack of conjugate priors for logistic models (Marin & Robert, 2007). However, it is straightforward to apply the random walk Metropolis Hastings approach, discussed in

Gelman et al. (2013), to simulate values from the posterior distribution. I propose the following implementation of this algorithm:

Iterate:

1. Generate an initial value for each element of $\theta$ from $U(-1,1)$, and initial state values $a_{1:PN}^{(0)}, w_{1:PN}^{(0)} \sim N(0,1)$.
2. Compute the likelihood function. Simulate the entire state paths, $a_{1:T}$ and $w_{1:T}$, for $NP$ particles from $a_{1:T}|\theta, w_{1:T}, x, y_{1:T}$, and $w_{1:T}|w_{0:T-1}$ via the propagate resample filter. Approximate the likelihood function as $\bar{L}(\theta) = \sum_i p(y_{1:T}|a_{1:T}, w_{0:T-1}, x, ; \theta)\ /NP$ by substituting the simulated vectors $a_{1:T}, w_{0:T-1}$; see Section 2.3.3 for details about particle sampling.
3. Metropolis-Hasting algorithm.
    a. Propose a new $\theta^* \sim N(\theta^s, \sigma^2 I)$ with $\sigma = .5$, which controls acceptance. Denote the proposal distribution $q(\theta^*, \theta^s)$.
    b. Compute the acceptance ratio,

$$r = \frac{L(\theta^*)P(\theta^*)q(\theta^s, \theta^*)}{L(\theta^s)\ P(\theta^s)q(\theta^*, \theta^s)} = \frac{L(\theta^*)P(\theta^*)}{L(\theta^s)\ P(\theta^s)},$$

    where $P(\theta) \sim N(0.\delta I)$ denotes an independent normal prior for the parameter vector and the proposal terms cancel due to the symmetry of the normal distribution.
    c. Let

$$\theta' = \begin{cases} \theta^* \ with \ probability \ \min(r, 1) \\ \theta^s \ with \ probability \ 1 - \min(r, 1) \end{cases}.$$

To achieve these proportions, sample $u \sim uniform(0,1)$ and set $\theta' = \theta^*$ if $u < r$ and $\theta' = \theta^s$ otherwise.

Because the state vector has a very simple time dependence structure, it is much more efficient to simulate from the posterior distribution of $\theta$ using particle filter techniques (Lopes & Tsay, 2011) than more traditional Markov Chain Monte Carlo techniques. Alternative approaches that integrate the particle filter with parameter estimation can be found in Petris, Campagnoli, and Petrone (2009) and Carvalho, Johannes, Lopes, and Polson (2010).

### 2.3.3 Sequential Monte Carlo and the Particle Filter

Although both maximum likelihood and Bayesian posterior sampling are feasible approaches for inference on the parameters in $\theta_i$, such methods are neither appropriate nor scalable for inferring the state vectors $a_t$ and $w_t$. Therefore, I approached this problem by deploying the two-stage simulation procedure discussed in the previous section. While the second, outer parameter learning stage, used a standard MCMC approach, the first stage used a recently introduced simulation approach designed for time sequence data. Although this approach has been used extensively for non-linear state space applications in general, it has only been experimented with briefly in the conditional logit or choice setting.

The particle filter is based on the traditional Monte Carlo integration technique known as importance sampling: a widely used fundamental technique that has been

extended in a wide variety of approaches (Fishman, 2005). Suppose that one is interested in computing a general integral of the form,

$$y = \int_a f(x)\, dx\,. \tag{2.3.6}$$

If the integral of $f(x)$ does not have an analytical solution, I may consider Monte Carlo methods; see Rizzo (2007) (Rizzo, 2007) for an introductory survey. While numerous approaches exist, one that may be particularly efficient in certain circumstances involves the equivalent form of the integral,

$$y = \int_a \frac{f(x)}{g(x)} g(x) dx. \tag{2.3.7}$$

A judicious choice of $g(x)$, one that is easy to generate samples from and that matches the shape and support of $f(x)$ well, will allow a high quality estimate of *y*. This is accomplished by sampling a series of independent observations $x_1, x_2, \ldots, x_K \sim g(x)$ and using them to compute the average,

$$\bar{y} = \frac{1}{K} \sum_{i=1}^{K} \frac{f(x_i)}{g(x_i)}. \tag{2.3.8}$$

This approach has been embellished and adapted to provide independent sampling in many contexts. Gordon (1993) (Gordon, Salmond, & Smith, 1993) and Doucet et al. (2000) (Doucet et al., 2000) first recognized the value of this approach in non-linear time series and filtering applications.

Returning to my algorithm, for purposes of exposition, I consider the analysis of repeated choices for a single individual. Importance sampling generates a set of simulated values and weights that can be used to approximate a distribution. The efficiency of the particle filter derives from the ability to simulate from the current distribution, $P(a_{1:t}|y_{1:t}, w_{0:t-1}, x;\ \theta)$, given the previous distribution, $P(a_{1:\mathrm{t}-1}|y_{1:t-1}, w_{0:t-2}, x;\ \theta)$, and the probability, $P(y_t|a_t;\ \theta)$. Instead of directly deriving the joint distribution with regard to particles at time $t$ and $t-1$, I consider the following relation:

$$P(a_t, a_{t-1}, w_t, w_{t-1}|y_t, y_{1:t-1}) \propto P(y_t|a_t) \cdot P(a_t|a_{t-1}, w_t) \cdot P(w_t|w_{t-1}) \cdot P(a_{t-1}|y_{1:t-1}).$$

Using this, draws can be obtained in two steps: 1) *recursively* propagate particles (simulate $a_t, w_t$) from the posterior at time $t-1$ to $t$, and 2) weight the particles proportionally based on their likelihoods; see Petris et al. (2009) and Lopes and Tsay (2011) for further details.

**Algorithm summary**

Draw $\left\{a_0^{(i)}\right\}_{i=1}^{NP}$ from $N(0,1)$, given $\left\{w_0^{(i)}\right\}_{i=1}^{NP}$.

For $i$ in $1{:}\ NP$

1. Propagate $\left\{w_{t-1}^{(i)}\right\}_{i=1}^{NP}$ to $\left\{\tilde{w}_t^{(i)}\right\}_{i=1}^{NP}$, $P(w_t|w_{t-1})$ via a random walk equation and propagate $\left\{a_{t-1}^{(i)}\right\}_{i=1}^{NP}$ to $\left\{\tilde{a}_t^{(i)}\right\}_{i=1}^{NP}$, $P(a_t|a_{t-1}, w_t)$ via a trend equation.
2. Resample $\left\{a_t^{(i)}\right\}_{i=1}^{NP}$ from $\left\{\tilde{a}_t^{(i)}\right\}_{i=1}^{NP}$ with weights $\delta_t^i \propto P\left(y_t \middle| \tilde{a}_t^{(i)}\right)$ and resample $\left\{w_t^{(i)}\right\}_{i=1}^{NP}$ from $\left\{\tilde{w}_t^{(i)}\right\}_{i=1}^{NP}$ with weights $\delta_t^i \propto P\left(a_t^{(i)} \middle| a_{t-1}^{(i)}, \tilde{w}_t^{(i)}\right)$.

I draw the initial particles $\left\{a_0^{(i)}\right\}_{i=1}^{NP}$ and $\left\{w_0^{(i)}\right\}_{i=1}^{NP}$ from a normal distribution, because the limiting distribution $\lim_{t\to\infty} p(a_t)$ of the latent state is again a normal distribution and, as the latent state evolves based on normal noise, the most natural distribution for the initial state is also the normal. Note that $t = 0$ does not indicate the first purchase of a customer, rather it is the first "observed" purchase of the customer. A key advantage of using state space models is that they have week dependency on the initial state distribution, which is discussed further in Section 2.4.1. As data on new purchases are added, the effect due to the initial state diminishes. On the other hand, if I use a nonlinear model with lagged dependent variables, resolving the initial conditions is more difficult (Wooldridge, 2005).

## 2.4 Advantages of Discrete Choice State Space Model

### 2.4.1 Accounting for Initial Conditions

As noted in Chapter 1, a common method of capturing state dependence in dynamic choice and other panel models is to use lagged versions of the response variable, see, for example, Guadagni and Little (1983). Regression models that use this formulation are usually called observed state dynamic models. When a regression

function for a categorical dependent variable contains lagged variables, I face the initial value problem (Hsiao, 2003). The challenge is that the model fit relies on responses that occur before the observation period begins. The lack of knowledge regarding this value leads to an estimator that is inconsistent and biased (Hsiao, 2003; Wooldridge, 2005). In the linear model case, an appropriate transformation can resolve this problem (Wooldridge, 2005).

In the dynamic discrete choice setting, the estimation of the lagged dependent model poses the difficulty of the appropriate treatment of the initial value of the outcome variable, $y_{i0}$. If $y_{i0}$ is strictly exogenous and fixed, then it can be modelled as an exogenous nonrandom condition similar to $x_{it}$. This assumption is typically problematic because independence between $y_{i0}$ and the vector $(x_{it}, a_i)$ is a very strong assumption. If $y_{i0}$ is treated as random, then the distribution takes the following form:

$$f(y_{i0}, y_{i1}, \dots, y_{iT}|x_{it}, a_i) = f(y_{i1}, \dots, y_{iT}|y_{i0}, x_{it}, a_i) \cdot f(y_{i0}|x_{it}, a_i). \quad (2.4.1)$$

To define the likelihood, $f(y_{i0}|x_{it}, a_i)$ needs to be specified, and this is difficult to know without significant information on the details of the specific process (Erdem & Sun, 2001; Kitamura & Bunch, 1990). Numerous authors have investigated the initial conditions problem. Heckman (1981a), Wooldridge (2005), and Honoré and Tamer (2006) all offer solutions of varying complexity. Miranda (2007) reviews methods to correct for bias caused by conditioning on the initial value of the outcome variable.

Although they find the method proposed by Heckman to be the most accurate across general cases, they simultaneously note the complexity of the correction.

In contrast, reiterating my earlier discussion in Section 2.2.2, the state space model captures correlation in the observed choices through a sequence of unobserved latent states $a_{it}$, which are correlated over time through a first-order Markov process, $a_{it} = \phi_i \cdot a_{i,t-1}$, where $t = 1, \dots, T$. Conditional on $x_{it}$ and the current state value, $a_{it}$, the current observation, $y_{it}$, is independent of all past and future values of both the state and outcome variables. By allowing the initial state, $a_{i0}$, to follow a distribution whose coefficients are independent of other observations, the uncertainty does not affect inference on the latent process or the exogenous variables (Heiss, 2008).

More concretely, the argument above says that

$$f\left(a_{it}|x_{it}, y_{i,t-1}, a_{i,1:t-1}; \phi_i\right) = f\left(a_{it}|x_{it}, a_{i,t-1}, \phi_i\right), \tag{2.4.2}$$

while

$$P\left(y_{it}|x_{it}, y_{i,1:t-1}, a_{i,1:T}\right) = P(y_{it}|x_{it}, a_{it}), \tag{2.4.3}$$

where $i = 1, \dots, N$ and $t = 1, \dots, T$. Working backward, I find $P(y_{i1}|y_{i0}, a_{i0};\ \theta_i) = P(y_{i1}|a_{i1}; \theta_i)P(a_{i1}|a_{i0}, y_{i0}; \theta_i)P(y_{i0}|a_{i0}; \theta_i)P(a_{i0})$ and, by allowing $a_{i0} \sim N(a, b)$, the initial value is correctly accounted for. Hence, the assumption of a latent habit persistence process, such as no feedback from prior outcome variables, avoids the usual initial value problem.

### 2.4.2 Misspecification in Serial Correlation

As alluded to in the previous section, a challenge of the traditional dynamic choice models with lagged variables, given in Equation 1.4, is that estimates of the coefficients $\beta$ are only consistent if the functional form for serial correlation in the error terms, $\epsilon_{ijt}$, is correctly specified. In addition, according to Heckman (1981a), one will get spurious state dependence, such as erroneously concluding that $\phi$ is significant, unless the serial correlation structure is properly specified. Erdem and Sun (2001) similarly note:

> *… if the serial-correlation structure is misspecified, the lagged dependent variables may be spuriously significant, simply because they help to fit the temporal dependency in the data better… For instance, if the errors are AR(1) (first order autoregressive) and the econometrician assumes random effects, it will also lead to inconsistent estimates of lagged dependent variable coefficients as will any misspecification of the serial correlation structure.*

Because state space dynamic choice models do not contain lagged dependent variables, they may be less sensitive to misspecification of the serial correlation structure. Erdem & Sun (2001) maintain, "As is well known, if the heterogeneity and/or serial-correlation structure is misspecified in models that do *not* contain lagged endogenous variables, it typically only causes inefficiency, not inconsistency of the estimates." G Chamberlain (1978) goes further and argues that non-linear panel models without lagged

dependent variables will be robust to the misspecification of errors; however, to my knowledge, there is no theoretical proof that parameter estimates from this model are insensitive to arbitrary specification of serial correlation in errors (Erdem & Sun, 2001). More recently, Keane (1997) and Erdem and Sun (2001) discuss previous empirical findings on discrete choice models with random intercepts that indicate the coefficients of the model are slightly effected by misspecification of serial correlation structure.

In Chapter 3, I explore this potential advantage through a simulation experiment. The experiment tests the sensitivity of the state space dynamic choice model to misspecification of the functional form for heterogeneity and serial correlation by fitting the proposed model to data generated from a more complex process.

## 2.5 Conclusion

This chapter introduced a new state space model for dynamic choice behavior that captures correlation in observed choice outcomes through a latent state term, $\boldsymbol{a}_{ijt}$. Using the results of Akaike (1974), I show that this state space model is equivalent to the lagged utility model proposed by Heckman (1981b). This model provides a valuable alternative approach to state dependence models that use lagged dependent variables to capture correlation.

The state space approach is important and allows us to implement the particle filtering to more efficiently compute the likelihood and perform inference. I integrate the particle filter with traditional MCMC methods to simulate posterior estimates of the parameters. Without a state space representation simulation, steps utilizing traditional

MCMC methods would be much more time consuming. In addition, the method facilitates fast updating, which is invaluable in real world applications where an operation may need to track thousands of products in a large number of categories.

Finally, I addressed two common problems in dynamic choice models based on state dependence terms, initial conditions, and misspecification in serial correlation. The state space dynamic choice is not impacted by initial conditions, since it does not require boundary values of the dependent variable. Since state space techniques easily decompose the random component of utility and are designed to capture correlation in these components over time, the flexibility of my model is an additional advantage in controlling for serial correlation.

# Chapter 3

## Simulation Studies of the State Space Discrete Choice Model

Chapter 2 introduced a new state space based choice model which was designed to capture habit persistence with a different modeling mechanism than had previously been considered. A new fitting algorithm was also proposed based on a sequential importance sampling technique called the particle filter. This chapter focuses on 1) testing and calibrating the algorithm for accuracy and convergence and 2) testing the algorithm under mild misspecification in order to assess convergence.

### 3.1 Testing Properly Specified Models

To begin I consider several simulation experiments in order to assess the effects of Monte Carlo particle filter sampling steps on the finite sample accuracy of the posterior estimates with varying sample sizes.

#### 3.1.1 Simulation Experiment 1

I start by considering a simple time series setting where I collect a sample of $T$ observations for a single individual. The goal of the first experiment is to explore the effect of both $T$, $N.P.$, the number of particles used in the particle filter to integrate over the state space terms, $a_{it}$, and $\phi_i$, the habit persistence term. I consider the effects of three factors on the fitting accuracy of data generated from the state space choice model given in Equations 3:1-3:3 below:

$$Y_{ijt}^{*} = a_{ijt} + w_{ijt}$$

$$a_{ijt} = \phi_i \cdot a_{ij,t-1} + \alpha_j + x_{ijt} \cdot \beta_i + w_{ijt}$$

$$w_{ijt} = \lambda_i \cdot w_{ij,t-1} + \zeta_{ijt} \quad \zeta_{ijt} \sim N(0, \sigma_\zeta)$$

For every combination of sample size, *T* = 15, 30, and 50, *N.P.* = 700, 3000, and $\phi_i = \{0.9, 06, 0.3, 0, -0.3, -0.6, -0.9\}$, one hundred individual time series were created by converting the final *T* observations of the state space simulation to a two state binary choice sequence by applying the logit transformation and classifying $Y_{it} = 1$ when $p_{i1t} \geq .5$ and $Y_{it} = 0$ when $p_{i1t} < .5$ (Shalizi, Forthcoming). In generating the data I set both $\beta_i = 0, \lambda_i = 0$ in above equations.

**Table 3.1 - Only Habit Persistent Model**

| $T$ | $S$ | $N.P$ | $\phi_i$ | $\widehat{\phi}_i$ | RMSE | Avg.Acc.R |
|---|---|---|---|---|---|---|
| 15 | 5000 | 700 | .9 | .74 | .278 | 9.2% |
| 15 | 5000 | 700 | .6 | .33 | .274 | 11.4% |
| 15 | 5000 | 700 | .3 | .26 | .139 | 12.8% |
| 15 | 5000 | 700 | .0 | .11 | .397 | 8.4% |
| 15 | 5000 | 700 | -.3 | -.10 | .423 | 12.4% |
| 15 | 5000 | 700 | -.6 | -.44 | .248 | 11.5% |
| 15 | 5000 | 700 | -.9 | -.75 | .275 | 9.5% |
| 15 | 5000 | 3000 | .9 | .88 | .185 | 28.7% |
| 15 | 5000 | 3000 | .6 | .63 | .176 | 27% |
| 15 | 5000 | 3000 | .3 | .41 | .433 | 35.3% |
| 15 | 5000 | 3000 | .0 | .06 | .287 | 35.8% |
| 15 | 5000 | 3000 | -.3 | -.27 | .215 | 25.4% |
| 15 | 5000 | 3000 | -.6 | -.57 | .230 | 34.6% |
| 15 | 5000 | 3000 | -.9 | -.93 | .202 | 25.7% |
| 30 | 5000 | 700 | .9 | .92 | .105 | 22.8% |
| 30 | 5000 | 700 | .6 | .66 | .236 | 27.2% |
| 30 | 5000 | 700 | .3 | .37 | .344 | 29.2% |
| 30 | 5000 | 700 | .0 | -.02 | .349 | 19.6% |
| 30 | 5000 | 700 | -.3 | -.36 | .261 | 29.5% |
| 30 | 5000 | 700 | -.6 | -.67 | .271 | 29.1% |
| 30 | 5000 | 700 | -.9 | -.91 | .088 | 27% |
| 30 | 5000 | 3000 | .9 | .89 | .049 | 24.3% |
| 30 | 5000 | 3000 | .6 | .63 | .213 | 19.3% |
| 30 | 5000 | 3000 | .3 | .34 | .147 | 10.2% |
| 30 | 5000 | 3000 | .0 | -.06 | .236 | 13.1% |
| 30 | 5000 | 3000 | -.3 | -.40 | .308 | 16.2% |
| 30 | 5000 | 3000 | -.6 | -.59 | .241 | 19.6% |
| 30 | 5000 | 3000 | -.9 | -.92 | .041 | 25.4% |
| 50 | 5000 | 700 | .9 | .80 | .175 | 1.6% |
| 50 | 5000 | 700 | .6 | .49 | .215 | 11.4% |
| 50 | 5000 | 700 | .3 | .19 | .203 | 9.5% |
| 50 | 5000 | 700 | .0 | .04 | .079 | 7.6% |
| 50 | 5000 | 700 | -.3 | -.27 | .094 | 8.3% |
| 50 | 5000 | 700 | -.6 | -.51 | .199 | 11.1% |
| 50 | 5000 | 700 | -.9 | -.83 | .109 | 1.2% |
| 50 | 5000 | 3000 | .9 | .91 | .053 | 7.8% |
| 50 | 5000 | 3000 | .6 | .56 | .115 | 14.6% |
| 50 | 5000 | 3000 | .3 | .26 | .143 | 12.1% |
| 50 | 5000 | 3000 | .0 | .03 | .079 | 11.9% |
| 50 | 5000 | 3000 | -.3 | -.27 | .124 | 13.3% |
| 50 | 5000 | 3000 | -.6 | -.54 | .179 | 16.8% |
| 50 | 5000 | 3000 | -.9 | -.88 | .079 | 7.2% |

Table 3.1 Continued.

| $T$ | $S$ | $N.P$ | $\phi_i$ | $\widehat{\phi}_i$ | RMSE | Avg.Acc.R |
|---|---|---|---|---|---|---|
| 100 | 5000 | 700 | .9 | .92 | .048 | 14% |
| 100 | 5000 | 700 | .6 | .63 | .051 | 23.9% |
| 100 | 5000 | 700 | .3 | .28 | .047 | 14.9% |
| 100 | 5000 | 700 | .0 | -.02 | .018 | 14.1% |
| 100 | 5000 | 700 | -.3 | -.29 | .031 | 18.4% |
| 100 | 5000 | 700 | -.6 | -.63 | .023 | 26.9% |
| 100 | 5000 | 700 | -.9 | -.92 | .030 | 16.3% |

Note: T, S, RMSE and N.P stand for number of transactions, iterations, root mean square of error, and particles, respectively. Avg.Acc.R indicates averaged acceptance rate in Metropolis-Hastings algorithm.

I then fit a model identical to the data generating process to each of the one hundred simulated data sets. In order to fit the model I applied the two-stage algorithm of Section 2.4. For each individual simulated data set the estimate of $\phi_i$ was based on a posterior sample of 5000 values. The first 3,000 iterations were used as a "burn-in" period, and the last 2,000 iterations were used to estimate the conditional posterior expectation and standard deviation. This procedure was repeated for each of the 100 data sets generated to assess the estimation accuracy. The results of the simulation are given in Table 3.1. The column labeled $\widehat{\phi}_i$ contains the mean of the 100 runs while Root Mean Square of Error, $RMSE = \sqrt{\left(\frac{\sum(\phi_i - \phi)^2}{100}\right)}$. The Avg.Acc.R provides information about the acceptance rate for $\phi_i$ in the Metropolis-Hastings chain. The acceptance rate depends on the proposal distribution; I used a normal-independence chain in this approach as discussed in Section 2.3. Asymptotic results suggest that an acceptance rate of 25% is optimal in producing the fastest possible convergence (Roberts, Gelman, & Gilks, 1997).

Table 3.1 shows that when using 3000 particles the algorithm provides significantly lower estimation error than when 700 are used, even in the case of a small

sample size such as 15 (time points). This improved performance comes at the cost of increased computing time. In this simple habit persistence only model it is clear that all procedures perform well for sample sizes greater than 30. Thus, the algorithm is less sensitive to the number of particles when the sample size is moderate to large, i.e. over 30. However, if I have small sample size like 15 observations, I can improve the estimation by increasing the number of particles. Furthermore, as the true value of $\phi_i$ moves away from the zero, the RMSE value decreases in most cases. This means that if strong habit persistence (inertia, or variety seeking) is manifested in the choice patterns, my procedures achieve improvement in estimation. In summary, I see that the proposed estimation method is effective in estimating the inertia parameter $\phi_i$ but that increasing the number of particles will improve accuracy, particularly in small samples.

### 3.1.2 Simulation Experiment 2

Under the same setup in previous section, my second experiment considers both habit persistence and price effects with parameter values, $\beta_i = -0.3, \lambda_i = 0$, and $\phi_i = (0.9, 06, 0.3, 0, -0.3, -0.6, -0.9)$. To test the impact of finite sample sizes I considered $T$ = 15, 30, 50 and 100 time points in this simulation.

**Table 3.2 - Habit Persistent and Price Effect Model**

| $T$ | $S$ | $N.P$ | $\phi_i$ | $\hat{\phi}_i$ (RMSE) | $\beta_i$ | $\widehat{\beta}_i$ (RMSE) | Avg. Acc.R |
|---|---|---|---|---|---|---|---|
| 15 | 2000 | 700 | .9 | .71 (.502) | -.3 | -.34 (.203) | 33.3% |
| 15 | 2000 | 700 | .6 | .45 (.526) | -.3 | -.19 (.251) | 32.1% |
| 15 | 2000 | 700 | .3 | .21 (.378) | -.3 | -.49 (.106) | 25.3% |
| 15 | 2000 | 700 | .0 | -.06 (.274) | -.3 | -.17 (.245) | 23.7% |
| 15 | 2000 | 700 | -.3 | -.39 (.117) | -.3 | -.33 (.221) | 39.4% |
| 15 | 2000 | 700 | -.6 | -.58 (.234) | -.3 | -.32 (.229) | 31.8% |
| 15 | 2000 | 700 | -.9 | -.77 (.409) | -.3 | -.12 (.206) | 31.5% |
| 15 | 2000 | 3000 | .9 | .91 (.284) | -.3 | -.22 (.128) | 32.8% |
| 15 | 2000 | 3000 | .6 | .47 (.411) | -.3 | -.23 (.095) | 34.8% |
| 15 | 2000 | 3000 | .3 | .23 (.367) | -.3 | -.24 (.087) | 35.4% |
| 15 | 2000 | 3000 | .0 | -.20 (.324) | -.3 | -.25 (.069) | 36.5% |
| 15 | 2000 | 3000 | -.3 | -.51 (.244) | -.3 | -.21 (.340) | 34% |
| 15 | 2000 | 3000 | -.6 | -.47 (.339) | -.3 | -.21 (.108) | 23.7% |
| 15 | 2000 | 3000 | -.9 | -.95 (.290) | -.3 | -.16 (.152) | 32.6% |
| 30 | 2000 | 700 | .9 | .81 (.359) | -.3 | -.17 (.122) | 23.6% |
| 30 | 2000 | 700 | .6 | .77 (.888) | -.3 | -.21 (.104) | 31.4% |
| 30 | 2000 | 700 | .3 | .23 (.339) | -.3 | -.25 (.055) | 28.3% |
| 30 | 2000 | 700 | .0 | -.04 (.400) | -.3 | -.24 (.063) | 27.8% |
| 30 | 2000 | 700 | -.3 | -.31 (.211) | -.3 | -.24 (.056) | 34.1% |
| 30 | 2000 | 700 | -.6 | -.45 (.285) | -.3 | -.23 (.069) | 31.1% |
| 30 | 2000 | 700 | -.9 | -.84 (.334) | -.3 | -.19 (.102) | 23% |
| 30 | 2000 | 3000 | .9 | .86 (.237) | -.3 | -.25 (.067) | 24.8% |
| 30 | 2000 | 3000 | .6 | .66 (.289) | -.3 | -.24 (.063) | 16.7% |
| 30 | 2000 | 3000 | .3 | .22 (.240) | -.3 | -.27 (.035) | 28.3% |
| 30 | 2000 | 3000 | .0 | -.02 (.238) | -.3 | -.23 (.064) | 27.8% |
| 30 | 2000 | 3000 | -.3 | -.21 (.292) | -.3 | -.32 (.023) | 27.5% |
| 30 | 2000 | 3000 | -.6 | -.55 (.241) | -.3 | -.31 (.023) | 28.6% |
| 30 | 2000 | 3000 | -.9 | -.88 (.174) | -.3 | -.24 (.058) | 25.6% |
| 50 | 2000 | 700 | .9 | .93 (.113) | -.3 | -.22 (.210) | 11.7% |
| 50 | 2000 | 700 | .6 | .45 (.364) | -.3 | -.26 (.102) | 19.2% |
| 50 | 2000 | 700 | .3 | .38 (.164) | -.3 | -.24 (.193) | 18.1% |
| 50 | 2000 | 700 | .0 | .12 (.141) | -.3 | -.18 (.207) | 21.77% |
| 50 | 2000 | 700 | -.3 | -.19 (.158) | -.3 | -.45 (.224) | 12.4% |
| 50 | 2000 | 700 | -.6 | -.57 (.130) | -.3 | -.32 (.192) | 16.3% |
| 50 | 2000 | 700 | -.9 | -.95 (.100) | -.3 | -.27 (.116) | 11.3% |
| 50 | 2000 | 3000 | .9 | .90 (.093) | -.3 | -.27 (.019) | 6.9% |
| 50 | 2000 | 3000 | .6 | .63 (.103) | -.3 | -.31 (.007) | 7.8% |
| 50 | 2000 | 3000 | .3 | .23 (.108) | -.3 | -.35 (.136) | 10% |
| 50 | 2000 | 3000 | .0 | -.01 (.017) | -.3 | -.30 (.008) | 13% |
| 50 | 2000 | 3000 | -.3 | -.25 (.095) | -.3 | -.26 (.105) | 9% |
| 50 | 2000 | 3000 | -.6 | -.57 (.102) | -.3 | -.36 (.010) | 11% |
| 50 | 2000 | 3000 | -.9 | -.91 (.052) | -.3 | -.25 (.106) | 7.3% |
| 100 | 2000 | 700 | .9 | .91 (.061) | -.3 | -.28 (.017) | 6.9% |
| 100 | 2000 | 700 | .6 | .63 (.098) | -.3 | -.32 (.019) | 7.8% |
| 100 | 2000 | 700 | .3 | .25 (.117) | -.3 | -.26 (.113) | 10% |

Table 3.2 Continued.

| $T$ | $S$ | $N.P$ | $\phi_i$ | $\hat{\phi}_i$ (RMSE) | $\beta_i$ | $\widehat{\beta_i}$ (RMSE) | Avg. Acc.R |
|---|---|---|---|---|---|---|---|
| 100 | 2000 | 700 | .0 | -.01 (.011) | -.3 | -.32 (.011) | 13% |
| 100 | 2000 | 700 | -.3 | -.26 (.074) | -.3 | -.27 (.095) | 9% |
| 100 | 2000 | 700 | -.6 | -.58 (.092) | -.3 | -.36 (.030) | 11% |
| 100 | 2000 | 700 | -.9 | -.89 (.053) | -.3 | -.26 (.093) | 7.3% |

Note: T, S, RMSE and N.P stand for number of transactions, iterations, root mean square of error, and particles, respectively. Avg.Acc.R indicates averaged acceptance rate in Metropolis-Hastings algorithm.

Inferences for $\phi_i$ and $\beta_i$ were based on 2000 samples from the posterior distribution. The first 1000 iterations were used as a "burn-in" period, and the last 1000 iterations were used to estimate the conditional posterior expectation and standard deviation. This procedure was repeated 100 times for a range of different $\phi_i$ values to assess the estimation accuracy.

The results of the simulation are given in Table 3.2. RMSE values indicate that using 3000 particles provides significantly lower estimation error than using 700 for sample sizes greater than 30 and that the proposed estimation method is capable of estimating the habit persistence (inertia) parameter $\phi_i$ and price coefficient $\beta_i$ quite accurately. I also find that increasing the number of particles does not have much impact on the accuracy of estimation in the case of small sample sizes such as 15 (time points) and that the estimates are much more variable in this case. Thus compared with previous simulation study in Table 3.1, adding one more parameter in my model requires a significant increase in sample size in order to provide accurate estimates.

I can also see clearly that the procedure is less sensitive to the number of particles when I have sample size greater than or equal to 50. Furthermore, if I have a moderate sample size like 30 transactions, I can improve the estimation by increasing the number

of particles. As in experiment 1 I also that as the true value of $\phi_i$ moves away from the zero, the RMSE value decreases in most cases. In summary, I see that the proposed estimation method is effective in estimating the habit persistent term $\phi_i$ and price coefficient $\beta_i$ but that including an additional parameter a larger sample size and increasing number of particles are necessary to achieve the same accuracy.

### 3.1.3 Simulation Experiment 3

I modifiy the data generating process in the previous experiment to include serial correlation in the error term in addition to habit persistence and a price effect. My parameterization $\beta_i = -0.3$, $\lambda_i = 0.3$, and $\phi_i = (0.9, 06, 0.3, 0, -0.3, -0.6, -0.9)$. In time series analysis, this type of model typically requires fairly large sample size (at least 50 time points is recommended by rule of thumbs). To test the performance of finite sample I considered 15, 30, 50 and 100 time points in this simulation.

**Table 3.3 - Habit Persistent, Price Effect, and Serial Correlation Model**

| $T$ | $S$ | $N.P$ | $\phi_i$ | $\hat{\phi}_i$ (RMSE) | $\beta_i$ | $\hat{\beta}_i$ (RMSE) | $\lambda_i$ | $\hat{\lambda}_i$ | Avg. Acc.R |
|---|---|---|---|---|---|---|---|---|---|
| 15 | 2000 | 700 | .9 | .58 (.511) | -.3 | -.24 (.176) | .3 | .28 (.561) | 27.8% |
| 15 | 2000 | 700 | .6 | .36 (.157) | -.3 | -.27 (.394) | .3 | .39 (.690) | 26.8% |
| 15 | 2000 | 700 | .3 | .16 (.400) | -.3 | -.29(.169) | .3 | .40 (.680) | 27.8% |
| 15 | 2000 | 700 | .0 | .15 (.499) | -.3 | -.31 (.172) | .3 | -.05(.824) | 28.3% |
| 15 | 2000 | 700 | -.3 | -.01 (.557) | -.3 | -.32 (.179) | .3 | -.02 (.948) | 27.9% |
| 15 | 2000 | 700 | -.6 | -.28 (.748) | -.3 | -.31 (.175) | .3 | -.02 (.805) | 27.7% |
| 15 | 2000 | 700 | -.9 | -.39 (.875) | -.3 | -.27 (.101) | .3 | -.17 (.903) | 28.7% |
| 15 | 2000 | 3000 | .9 | .46 (.447) | -.3 | -.24 (.183) | .3 | .39 (.489) | 27.6% |
| 15 | 2000 | 3000 | .6 | .39 (.413) | -.3 | -.26 (.187) | .3 | .31 (.562) | 28.3% |
| 15 | 2000 | 3000 | .3 | .19 (.492) | -.3 | -.29 (.181) | .3 | .18 (.595) | 27.5% |
| 15 | 2000 | 3000 | .0 | -.06 (.615) | -.3 | -.29 (.177) | .3 | .33 (.759) | 28.3% |
| 15 | 2000 | 3000 | -.3 | -.02 (.629) | -.3 | -.27 (.178) | .3 | -.10 (.881) | 28.0% |
| 15 | 2000 | 3000 | -.6 | -.23 (.688) | -.3 | -.32 (.169) | .3 | -.13 (.902) | 27.8% |
| 15 | 2000 | 3000 | -.9 | -.70 (.960) | -.3 | -.32 (.156) | .3 | .29 (.601) | 27.3% |
| 30 | 2000 | 700 | .9 | .63 (.423) | .9 | -.28 (.160) | .3 | .22 (.522) | 27.7% |
| 30 | 2000 | 700 | .6 | .39 (.355) | .6 | -.29 (.141) | .3 | .23 (.447) | 29.0% |
| 30 | 2000 | 700 | .3 | .26 (.402) | .3 | -.32 (.156) | .3 | .21 (.497) | 28.5% |
| 30 | 2000 | 700 | .0 | .17 (.445) | .0 | -.33 (.153) | .3 | .05 (.790) | 18.4% |
| 30 | 2000 | 700 | -.3 | -.07 (.598) | -.3 | -.35 (.176) | .3 | .13 (.650) | 27.8% |
| 30 | 2000 | 700 | -.6 | -31 (.852) | -.6 | -.33 (.163) | .3 | .19 (.542) | 28.1% |
| 30 | 2000 | 700 | -.9 | -.56 (.992) | -.9 | -.26 (.153) | .3 | .26 (.67) | 27.1% |
| 30 | 2000 | 3000 | .9 | .74 (.893) | -.3 | -.27 (.198) | .3 | .33 (.476) | 24.1% |
| 30 | 2000 | 3000 | .6 | .45 (.390) | -.3 | -.29 (.154) | .3 | .18 (.482) | 28.4% |
| 30 | 2000 | 3000 | .3 | .23 (.424) | -.3 | -.31 (.162) | .3 | .23 (.551) | 27.6% |
| 30 | 2000 | 3000 | .0 | -.04 (.617) | -.3 | -.33 (.156) | .3 | .27 (.275) | 27.5% |
| 30 | 2000 | 3000 | -.3 | -25 (.583) | -.3 | -.33 (.155) | .3 | -.21 (.898) | 28.1% |
| 30 | 2000 | 3000 | -.6 | -42 (.529) | -.3 | -.32 (.145) | .3 | -.12 (.772) | 28.5% |
| 30 | 2000 | 3000 | -.9 | -.62 (.905) | -.3 | -.30 (.106) | .3 | .22 (.650) | 27.9% |
| 50 | 2000 | 700 | .9 | .83 (.470) | -.3 | -.28 (.111) | .3 | .23 (.389) | 25.7% |
| 50 | 2000 | 700 | .6 | .64 (.446) | -.3 | -.31 (.073) | .3 | .13 (.576) | 29.1% |
| 50 | 2000 | 700 | .3 | .36 (.428) | -.3 | -.31 (.058) | .3 | .16 (.508) | 28.7% |
| 50 | 2000 | 700 | .0 | -.13 (.613) | -.3 | -.34 (.155) | .3 | .18 (.586) | 28.1% |
| 50 | 2000 | 700 | -.3 | -.33 (.469) | -.3 | -.31 (.094) | .3 | .13 (.582) | 29.0% |
| 50 | 2000 | 700 | -.6 | -.47 (.902) | -.3 | -.26 (.156) | .3 | .25 (.400) | 27.8% |
| 50 | 2000 | 700 | -.9 | -.86 (.677) | -.3 | -.24 (.160) | .3 | .23 (.354) | 26.1% |
| 50 | 2000 | 3000 | .9 | .87 (.374) | -.3 | -.27 (.109) | .3 | .24 (.313) | 29.9% |
| 50 | 2000 | 3000 | .6 | .63 (.479) | -.3 | -.32 (.069) | .3 | .19 (.438) | 25.6% |
| 50 | 2000 | 3000 | .3 | .34 (.361) | -.3 | -.33 (.109) | .3 | .28 (.391) | 29.1% |
| 50 | 2000 | 3000 | .0 | -.09 (.258) | -.3 | -.29 (.065) | .3 | .24 (.373) | 26.4% |
| 50 | 2000 | 3000 | -.3 | -.34 (.317) | -.3 | -.33 (.074) | .3 | .34 (.371) | 28.7% |
| 50 | 2000 | 3000 | -.6 | -.53 (.293) | -.3 | -.27 (.116) | .3 | .24 (.347) | 21.3% |
| 50 | 2000 | 3000 | -.9 | -.88 (.384) | -.3 | -.25 (.138) | .3 | .28 (.213) | 22.1% |
| 100 | 2000 | 700 | .9 | .92 (.325) | -.3 | -.28 (.062) | .3 | .21 (.281) | 24.3% |

Table 3.3 Continued.

| $T$ | $S$ | $N.P$ | $\phi_i$ | $\hat{\phi}_i$ (RMSE) | $\beta_i$ | $\hat{\beta}_i$ (RMSE) | $\lambda_i$ | $\hat{\lambda}_i$ | Avg. Acc.R |
|---|---|---|---|---|---|---|---|---|---|
| 100 | 2000 | 700 | .6 | .68 (.501) | -.3 | -.27 (.057) | .3 | .25 (.299) | 31.2% |
| 100 | 2000 | 700 | .3 | .37 (.329) | -.3 | -.31 (.044) | .3 | .19 (.422) | 29.2% |
| 100 | 2000 | 700 | .0 | -.07 (.151) | -.3 | -.34 (.102) | .3 | .23 (.572) | 29.6% |
| 100 | 2000 | 700 | -.3 | -.29 (.285) | -.3 | -.34 (.064) | .3 | .27(.283) | 28.9% |
| 100 | 2000 | 700 | -.6 | -.55 (.578) | -.3 | -.26 (.134) | .3 | .37 (.347) | 29.9% |
| 100 | 2000 | 700 | -.9 | -.96 (.625) | -.3 | -.25 (.145) | .3 | .36 (.202) | 23.0% |

Note: T, S, RMSE and N.P stand for number of transactions, iterations, root mean square of error, and particles, respectively. Avg.Acc.R indicates averaged acceptance rate in Metropolis-Hastings algorithm.

Again I generated 2000 values of the inertia parameter $\phi_i$, $\beta_i$, and $\lambda_i$ from the posterior distribution. The first 1000 iterations were used as a "burn-in" period, and the last 1000 iterations were used to estimate the conditional posterior expectation and standard deviation. This procedure was repeated 100 times for a range of different $\phi_i$, $\beta_i$, and $\lambda_i$ values to assess the estimation accuracy. The results of the simulation are given in Table 3.3.

Table 3.3 shows that increasing the number of particles does not improve estimation accuracy when the number of transactions (time points) is less than 30. When I have 50 transactions, I find that when using 3000 particles the procedures provides significantly lower estimation error than when using 700 particles and the proposed estimation method is capable of estimating the inertia parameter $\phi$, $\beta$, and $\lambda_i$ quite accurately. Unlike previous models, this data set with habit persistence, price effect, and serial correlation requires at least 3000 particles at sample size 50 in order to achieve reasonable accuracy. For sample sizes of 100, 700 particles are sufficient.

Unlike the first two experiments, as the true value of $\phi_i$ moves away from the zero, the RMSE value did not decreases in most cases. This means that even though

strong habit persistence (inertia, or variety seeking) are manifested in the choice patterns, my procedures could not achieve improvement in estimation when the model includes the serial correlation. In summary, I see that the proposed estimation method is effective in estimating the inertia parameter $\phi_i$, price effect $\beta_i$, and serial correlation $\lambda_i$ and both larger sample sizes and numbers of particles are critical.

### 3.2 Model Robustness against Misspecification of the Error Structure

Building on the discussion of Section 2.4.2 this section explores the impact of model misspecification in state space dynamic choice models. If habit persistence alone encapsulated the empirical reality in the case of repeated purchase, there would be no problem. However, Heckman (1981a) indicates that there is another possibility that could lead to spurious results when repeated purchases occurs because of unobserved factors. If such unobservable effects were systematic for same unit over transaction (or time), it could lead to a serial correlation in the error terms for those observations. Fitting such a model without accounting for errors would consistent but inefficient coefficient estimates, rendering any statistical testing inaccurate (Gulati & Gargiulo, 1999). If I fail to specify the correct order of serial correlation in the unobserved component, this improper treatment can lead to spurious effects appearing with attempts to assess the influence of previous utility on current decisions. For example, if the random components of utility function are first order autoregressive and I assume independence in errors, it will lead to inconsistent estimates for lagged dependent term because of misspecification of the serial correlation structure. However, because this type of misspecification

typically causes inefficiency not inconsistency in models that do not include lagged dependent variables I expect the state space dynamic choice model to perform reasonably well in this setting (Erdem & Sun 2001).

Building on earlier simulation experiments, I investigate robustness of parameter estimates from a habit persistence and price effect model with first order autoregressive structure when the true errors follow a second order autoregression. The data generating process is defined by the following equations:

$$Y_{ijt}^{*} = a_{ijt} + w_{ijt}$$

$$a_{ijt} = \phi_i \cdot a_{ij,t-1} + \alpha_j + w_{ijt} + x_{ijt} \cdot \beta_i$$

$$w_{ijt} = \lambda_i^1 \cdot w_{ij,t-1} + \lambda_i^2 \cdot w_{ij,t-2} + \zeta_{ijt}, \;\; \zeta_{ijt} \sim N(0, \sigma_\zeta)$$

This model allows us to evaluate the impact of misspecification on both coefficients for the habit persistent term, $\phi_i$ and price variable, $\beta_i$. I generate the data from the second order correlation by setting the parameter values: $\beta_i$ = -0.3, $\lambda_i^1 = 0.3$, $\lambda_i^2 = (-0.6, -0.3, 0.3, 0.6)$, $\phi_i = 0$ and $y = (0,1)$ and fit the model with first order correlation, $\lambda_i^2 = 0$. Procedures were performed using the two-stage estimation algorithm discussed in Section 2.3.2.

**Table 3.4 - Misspecification Check**

| $T$ | $S$ | $N.P$ | $\widehat{\phi_i}$(RMSE) | $\widehat{\beta_i}$(RMSE) | $\hat{\lambda}_i^1$(RMSE) | $\{\phi_i, \beta_i, \lambda_i^2\}$ | Avg. Acc.R |
|---|---|---|---|---|---|---|---|
| 50 | 300 | 3000 | .03 (.361) | -.23 (.311) | .14 (.254) | {0, -.3, -.6} | 17.4% |
| 50 | 300 | 3000 | -.09 (.117) | -.31 (.103) | .22 (.243) | {0, -.3, -.3} | 19.3% |
| 50 | 300 | 3000 | -.009 (.151) | -.26 (.217) | .17 (.273) | {0, -.3, .3} | 16.9% |
| 50 | 300 | 3000 | .08 (.463) | -.21 (375) | -.12 (.681) | {0, -.3, .6} | 24.4% |
| 100 | 300 | 3000 | .04 (.322) | -.23 (.041) | .17 (.353) | {0, -.3, -.6} | 27.8% |
| 100 | 300 | 3000 | -.02 (.101) | -.21 (.031) | .25 (.128) | {0, -.3, -.3} | 27.4% |
| 100 | 300 | 3000 | -.0.02 (.129) | -.23 (.043) | .15 (.288) | {0, -.3, .3} | 27.8% |
| 100 | 300 | 3000 | .06 (.420) | -.20 (.036) | -.13 (.406) | {0, -.3, .6} | 26.4% |

Note: T, S, and N.P stand for number of transactions, iterations, and particles, respectively. Avg.Acc.R indicates averaged acceptance rate in Metropolis-Hastings algorithm.

I generated 500 values of the inertia parameter $\phi_i$, the coefficient of the independent variable, $\beta_i$, and the first order serial correlation, $\lambda_i$, from the posterior distribution. The first 200 iterations were used as a "burn-in" period, and the last 300 iterations were used to estimate the conditional posterior expectation and standard deviation. This procedure was repeated 100 times for a range of different $\lambda_i^2$ values to assess the estimation accuracy. The results of the simulation are given in Table 3.4. The columns labeled $\hat{\phi}_i$, $\hat{\beta}_i$ and $\hat{\lambda}_i^1$ contain the means of the 50 runs while Root MSE represents the square root of the mean squared error.

Table 3.4 shows that robustness to misspecification of serial correlation depends on the magnitude of coefficient of second order serial correlation. When $\lambda_i^2 \in (-.3, +.3)$, the RMSE value of the habit persistence term, $\hat{\phi}_i$ is relatively lower than the case of $\lambda_i^2 \in (-.6, +.6)$. However, all $\hat{\phi}_i$ are very close to the true parameter value 0. Thus, I see that the estimate of the habit persistence coefficient is robust for the misspecification in serial correlation.

Unfortunately, $\hat{\beta}_i$ and $\hat{\lambda}_i^1$ tend to underestimate the true parameter values in most cases( an exception occurs when sample size is 50 and $\lambda_i^2 = -.3$. When the errors are highly correlated, i.e. $\lambda_i^2 = (-.6, +.6)$, the RMSE values of $\hat{\beta}_i$ and $\hat{\lambda}_i^1$ are larger than when moderate second order correlation exists. One way to think about this effect is that the effective degrees of freedom are far fewer than the number of observations because the residuals are more redundant (i.e., not independent one another) than the case of $\lambda_i^2 = (-.3, +.3)$.

Unfortunately, increasing sample size from *T*=50 to 100 does not seem to decrease the bias in the overall estimation when the model is misspecified with respect to serial correlation. In summary, I see that the under this data generating process, the proposed estimation procedure can accurately recover the coefficient for habit persistence but not the other terms. Furthermore, increasing the sample size in the range considered will not improve accuracy in this situation.

# Chapter 4

# Habit Persistence and State Dependence

## 4.1 Introduction

Why do some customers repurchase the same product regularly while others switch frequently, and how do these behaviors change across categories? In this chapter, I take a more detailed look at the two mechanisms proposed to explain repeated purchasing, *state dependence* and *habit persistence,* and how they have been operationalized in previous literature as well as the current work. What does each approach imply about the data-generating process and therefore is most appropriate for a given situation? While both approaches attempt to capture the observed patterns of repeat purchases, the previous discussion suggests that dependence upon only previous purchases is insufficient, and including a habit persistence term may produce a much richer model due to its autoregressive nature.

In the current chapter, I review and contrast these two processes, discuss their implementation in the extant literature, and study which approach, if either, is more sensible in analyzing repeat choice behavior in the context of fast-moving consumer goods. I also compare models based on state dependence and habit persistence through two case studies that investigate repeat purchases in fast-moving consumer goods. The first case study compares model fit on a data set capturing repeat purchases of pancake mix over a two-year period. Beyond comparing modelling approaches, this case study demonstrates the wide variation in repeat purchase propensity across the population as well as tremendous heterogeneity in price sensitivity across customers. The second case

study looks at habit persistence across categories of hedonic and utilitarian goods. The previous theory suggests that habit persistence should be weaker in hedonic product groups, and I investigate this hypothesis.

The results of this chapter are important because the simultaneous analysis of drivers of repeat purchase behavior and price sensitivity are critical factors in evaluating promotions and other pricing strategies. In addition, with the huge growth in the field of customer relationship management (CRM), understanding repeat purchasing behavior and the factors that influence it has become an issue of critical importance. The eventual goal is the ability to fashion programs that increase the frequency or consistency of purchases or widen the scope of the consumer's interaction with the firm and its partners (Venkatesan & Farris, 2012).

### 4.2. Habit Persistence

A habit originates as a performed activity that requires effort but after frequent repetition becomes automatic (Banerjee, 1994). Hence, after an initial period of feedback, the habitual behavior is no longer explained by the process of updating through trial and error in everyday experience (feedback). Instead, it is a formulated latent construct that controls the sequence of choice. Habit formation and habit persistence are widely referenced concepts in the economics literature; see Constantinides (1990). In economic terms, habit persistence is an economic term and refers to correlation in the latent utility of a choice or decision over time, which agrees with the above definition (Constantinides 1990, Seetharaman 2004, Heckman 1981b). Also consistent with this definition, habitual

purchasing may be considered rational behavior because it helps consumers to achieve satisfactory outcomes (maximize utility) by minimizing the costs of thinking and simplifying the decision-making process (Corstjens & Lal, 2000).

According to the Food Marketing Institute (www.fmi.org), the average American supermarket carries 42,686 items in 2012, which is more than five times the number in 1975. Britain's Tesco stocks 91 different shampoos, 93 varieties of toothpaste, and 115 types of household cleaners. Theoretical and empirical evidence suggest that habitual purchasing plays an important role in these low-involvement environments (Corstjens & Lal, 2000).

Similar to our description of habitual purchasing, Mellens et al. (1996) (Mellens, 1996) define inertia as the propensity for consumers to stay with the same brand because they are not prepared to spend effort and time to search for other brands. The implication is that consumers are using prior utility to form current evaluations, as in habit persistence. In the research that follows, I will use habit persistence and inertia synonymously to indicate a dependence over time for choice utilities.

Bawa (1990), also working in a choice model context, introduced a quadratic model of utility based on a count of the prior number of purchases of the same product since the last product switch. The linear and quadratic parameters of this model were then used to define a range of four categories of variety-seeking and inertial behavior. A potential weakness of this model, as the author notes, is that inertia here depends strictly on the length of the current run of purchases. So, if a person purchases the product on numerous prior occasions but made a single switch on the previous purchase, then his

inertia state reverts to 0. In effect, the model has no memory of the previous purchase. However, a number of factors could cause single-brand switches, like a change in shopper, stock-outs, or an inability to find the appropriate product, and should not totally negate the effect of previous purchase history. As a result, despite the long history of work in this field, there is ample room for improved forms of modelling that provide more intuitive and accurate measures of inertia and allow us to estimate the impact of programs and incentives on inertial shopping patterns.

In the context of labor economics, Heckman (1981b) proposes to operationalize habit persistence through a lagged utility term. Although possibly complementary, this approach obtains a dynamic utility function which depends explicitly on previous utilities, as opposed to capturing this effect indirectly through lagged outcomes of previous periods. Heckman proposes a general mathematical term to operationalize habit persistence in panel models. Define the current relative utility, $Y_{it}^*$, then

$$Y_{it}^* = G(L)Y_{it}^* + \epsilon_{it}, \qquad (4.1)$$

where $G(0) = 0$ and $G(L)$ is a general (or polynomial) lag operator of order $K$, [$G(L) = g_1L + g_2L^2 + \cdots + g_KL^K,\ where\ L^KY_{it}^* = Y_{i(t-k)}^*$]. This term describes the cumulative effect of *previous memory of utility* on current choices. In the first order case,

$$Y_{it}^* = g_1Y_{it-1}^* + \epsilon_{it},$$

it is easy, through recursive substitution, to see that the model is equivalent to $Y_{it}^* = \sum_{k=0}^{\infty} g_1^k \epsilon_{i(t-k)}$. Hence, allowing current utility to be a function of lagged utility implies

that utility is actually an exponentially smoothed function of the entire history of the utility process. Whether or not such an approach is most appropriate will be context dependent, but the model is much richer in using the process history than it initially appears. This approach was advocated in the much earlier "latent Markov" model (Coleman, 1964), in which prior propensities to select a state rather than prior occupancy of a state determine the current probability that a state is occupied.

In the context of retail shopping, variety seeking is defined by the utility the consumer derives from the change in a choice itself, irrespective of the brand she switches to or from (Seetharaman et al. 1998) (Seetharaman & Chintagunta, 1998). Because variety seeking is driven by changes in utility, it is logical to model variety seeking as a latent utility process. I argue that the state space dynamic choice model, Equation 2.2.13, can naturally capture variety-seeking behavior using the same mechanism that captures habit persistence. The coefficient of the habit persistence term captures the dynamic tendencies of purchasing behavior. The sign of this parameter reveals whether individuals are inertial (+) or variety seeking (-) in nature. The negative estimated value would then imply that the second-highest utility product in the previous choice is more likely to be selected in the current choice, which is consistent with a variety-seeking explanation.

### 4.3. State Dependence

In a wide range of social science research, such as labor force participation, the incidence of accidents, and unemployment, it is known that individuals who have

experienced an event in the past are more likely to experience the event in the future than are individuals who have not experienced the event (Heckman 1981a, Bates and Neyman 1951, Layton 1978, Heckman and Willis 1977) (Bates & Neyman, 1952; Heckman, 1981a; Heckman & Willis, 1977; Layton, 1978). It follows that models describing this behavior should include features that allow current event probabilities to be a function of previous events, i.e., current probabilities should differ based on the individual's history.

The idea that historical decisions may shape a decision maker's current preferences follows naturally from common sense and personal experience. Heckman (1981a) notes, "… 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. Structural relationships of this sort give rise to true state dependence …" He formalizes this empirical regularity in general panel models by including a mechanism to capture the concept of *state dependence*.

Heckman (1981a), writing in the context of labor economics, proposed the following expression in order to capture the effect of previous events on the present. The utility for individual $i$, $Y_{it}^{*}$, can be approximated by

$$Y_{it}^{*} = X_{it}\beta + \sum_{t'<t} \delta_{t,t}\, d_{it'} + \sum_{j} \lambda_{t,(t-j)} \prod_{l=1}^{j} d_{i(t-l)} + \varepsilon_{it} \qquad (4.2)$$

where $i = 1, \ldots, I$ and $t = 1, \ldots T$. $E(\varepsilon_{it}) = 0$, $E(\varepsilon_{it}, \varepsilon_{it\prime}) = \sigma_{t,t'}$. $E(\varepsilon_{it}, \varepsilon_{it\prime}) = 0$, $i \neq i'$. $X_{it}$ is a vector of exogenous variables that determine choices in period $t$. $\beta$ is a suitably dimensioned vector of coefficients. The effects of previous work experience on choice in period $t$ are captured by the second and third terms on the right-hand side of the equation 4.2.

The second term, $\sum_{t'<t} \delta_{t,t\prime} d_{t,t\prime}$, indicates the effect of all prior experience on choice in period $t$. The third term, $\sum_j \lambda_{t,t-j} \prod_{l=1}^{j} d_{i,t-l}$, specifies the effect on choice of experience in period $t$ in the most recent continuous spell of work for those who have worked in period $t-1$. The coefficients associated with these terms are written to allow for depreciation of the effect of prior work experience. Setting $\delta_{t,t\prime} = \delta(t - t')$ for $t - t' \leq K$, $\delta_{t,t\prime} = 0$ otherwise generates a $K$th order Markov process. Heckman records whether or not individual $i$ works at time $t$ by introducing a dummy variable $d_{it}$ that assumes the value of one when the individual works at that time, and zero otherwise. As in a standard probit approach, $d_{it} = 1$ if $Y_{it}^* > 0$, while $d_{it} = 0$ if $Y_{it}^* \leq 0$.

Figure 4.1 provides a simple schematic explanation showing how current behavior directly impacts future utility. In Chapter 1, I described such a process, in the context of choice outcomes, as a discrete state Markov Chain. This follows because the current state is defined by the current choice category, which comes from a finite set, and this choice defines the current utilities and consequently defines the subsequent choice probabilities.

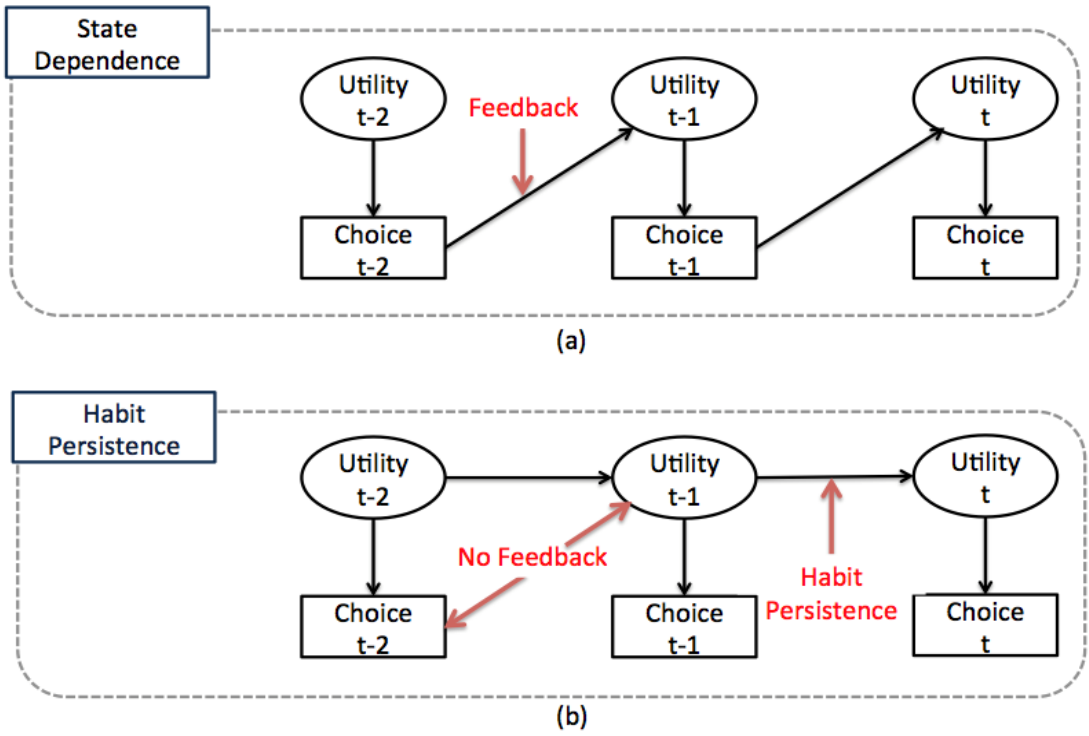



**Figure 4.1 - Habit Persistence and State Dependence**

In contrast, as discussed in Section 4.2, a pure habit persistence process is characterized by no direct feedback from earlier events. Instead, all correlation between events is based on autocorrelations between the underlying utility. An AR(1) version of this process is depicted in the lower panel of Figure 4.1. Instead of a discrete state Markov process, I have a continuous autoregressive process defining the choice probabilities. Such a process may be viewed as governed by external experiences, such as advertising, packaging, prior opinion, and other influences outside of direct prior consumption.

In order to control for heterogeneity across customers in an early discrete choice modelling framework, Guadagni and Little (1983) defined an exponential smoothing of the binary sequence of yes/no purchases, a term that is now known as the GL – index. (See Section 1.2 for a detailed description.) Despite the original goal of using this index

to control for heterogeneity, this measure became a de facto operational measure of loyalty; see Corstjens and Lal (2000). Although the loyalty interpretation remains controversial (McAlister 1991) (McAlister et al., 1991), the term is still widely used to capture heterogeneity and state dependence in choice models (Keane 2013).

### 4.4 State Dependence, Habit Persistence, and Serial Correlation in Models of Supermarket Scanner Data

A common context for the application of dynamic choice models is the analysis of purchases of fast-moving consumer goods recorded in supermarket scanner data. Guadagni and Little (1983) were the first to apply the choice model of McFadden (1974) to the analysis of retail grocery data. In the 30 years that followed, this topic was revisited frequently in several streams of literature.

Keane (1997, 2013), along with colleagues, extended this earlier work using a GL term to capture state dependence while including correlated error terms to protect against inconsistency of estimates and potential spurious state dependence. In contrast to logit-based models, the probit approach also allows other features, such as associations between choices, to be modeled in more detail. Along with colleagues, Keane also developed the widely used GHK simulation technique to fit the proposed dynamic choice probit model. This approach is then applied to fast-moving consumer goods categories with the goal of understanding the net impact of promotion strategies based on price. The presence of state dependence terms leads to the prediction that most gains, because of

price promotion, are due to cannibalization of future sales and short-term switching behavior. This model is summarized in column 3 of Table 4.1.

Allenby & Lenk (1994) focus instead on a logit formulation of the model, which captures neither habit persistence nor state dependence, but simply focuses on the effects of choice features while controlling for serial correlation. This model uses a logit formulation, implying extreme value random errors, but simultaneously captures autocorrelation in error terms within the logit equation using normal error structures. Although not directly discussed, this indicates an implicit partitioning of the error into two distinct pieces. I used a similar error structure in the experiments presented in Chapter 3 and the case studies presented in Section 4.5.

A unique set of literature that deviates in a creative and important way from the models discussed above uses the *Lightning Bolt* (LB) formulation as described in Roy et al. (1996), Chintagunta (1998, 1999), and Seetharaman (2004) in order to capture habit persistence and state dependence. In the initial work, Roy et al. (1996) divide the utility into two components. The first is a fixed component that is a function of observables, like product features, price, and lagged observations. The second component is a random term that summarizes the inflow of information to the individual household decision maker and can be viewed as an error term, or shock. The novel aspect of the LB approach is that the current value of the error terms is not an MA(1) process, but instead follows an extreme process where the current value of the random term is the maximum of the most recent new error and the largest previous error. This process models an information flow where the random component depends not on a weighted average of

previous information – such as advertising, word of mouth, or other stimuli – but, instead, on the most impactful single stimuli. The random component of utility reflects the single most impactful, unmeasured information relating to a specific brand. Based on their model assumptions, this error structure leads to a closed-form expression for the correlation between probability transitions; see column 2 of Table 4.1. Arguably, this model contains both a state dependence term through the fixed utility term and a habit persistence term via the maximal error process. Chintagunta (1999) extends this model to handle variety-seeking behavior, while Seetharaman (2004) attempts to extend the model to include an additional habit persistence term using lagged utility.

The concepts of habit persistence and state dependence are distinct and clearly operationalized in Equations 4.1-2. Furthermore, these processes may function uniquely or in combination in a choice process. However, on balance, I find much more confusion and variability in the use of the term habit persistence (inertia) as well as in the modelling structures used to operationalize this concept. With the exception of Allenby & Lenk (1994), all of the models referenced in this section include one or more terms designed to capture state dependence by using a lagged dependent variable structure (the second term of Equation 4.2). Alternatively, habit persistence is either equated with state dependence as suggested in Keane (2013) or assumed to arise from autocorrelation in errors, as in Keane (1997, 2013) and Roy et al (1996). It is only (Haaijer & Wedel, 2001) and Seetharaman (2004) that return to Heckman's definition of state dependence, but it is notable that Seetharaman eventually deviates substantially from it.

While the reasons for this imbalance in coverage are unclear, and may be due to the practical issues of fitting nonlinear latent state models in earlier decades with less computing resources, it is clear that neither process can be assumed to take precedence in general. Additionally, habit persistence may play an important role in the purchase process of FMCG.

The previous discussion allows us to highlight several key contributions of the current work to the literature on dynamic choice and particularly in the context of scanner data. First, the proposed state space choice model provides the flexibility to capture habit persistence as well as state dependence in a very simple compact structure that is both easily identified and equivalent to the structure proposed by Heckman (1981a, 1981b). Second, the proposed model-fitting procedure, based on the particle filter, allows a wide range of models to be fit, including state dependence only, habit persistence only, and joint models, and captures autocorrelation in the error term. This flexibility allows both standard- and Bayesian-model selection procedures to be used to identify which processes are most critical in the particular context. The Bayesian approach also simplifies forecasting and the computation of important quantities, such as marginal effects.

**Table 4.1 - Summary of Discrete Choice Models**

| | **Allenby and Lenk's logistic normal regression model (Allenby and Lenk, 1994)** | **Roy, Chintagunta, and Haldar's suggested model (Roy et al. 1996)** | **Keane's typical structure of panel data discrete choice models (Keane 2013)** |
|---|---|---|---|
| Modelling equation | $y_{it}(j) = [\alpha_o(j) + \beta_{io}(j)] + x_{it}(j)'[\alpha_1 + \beta_{i1}] + d_{it}'\alpha_2(j) + \epsilon_{it}(j)$ | $P_t[j] = \frac{e^{v_t^{(j)}}}{\sum_{r=1}^{d} e^{v_t^{(r)}}}$, where $v_t^{(j)}$ is the systematic component of utility for brand $j$, and there are $d$ brands in the choice set. $v_t^{(j)} = \alpha^{(j)} + x^{(j)} \cdot \Theta + I_{t-1}^{(j)} \cdot \theta_{l+1}$. | $U_{ijt} = \alpha_{ij} + X_{ijt}\beta + \gamma d_{ijt}$ |
| Lagged dependent variables | n.a. | $I^{(j)} \cdot \theta_{l+1}$, the influence of observed past experience is accommodated by lagged choice variables. | $d_{ijt} = 1\ if\ U_{ijt} > U_{ikt}$ $= 0\ otherwise$ |
| Error terms that follow MA(1) process | $\epsilon_{it} = \Phi\epsilon_{i,t-1} + \zeta_{it}$ | | $\epsilon_{ijt} = \rho\epsilon_{ij,t-1} + \eta_{ijt}$ |
| Correlation between alternatives | | $P_{st}[i\|w]] = \begin{cases} (1-\rho)\frac{e^{v_t^{(j)}}}{\sum_{r=1}^{d} e^{v_t^{(r)}}} & if\ i \neq w, \\ (1-\rho)\frac{e^{v_t^{(j)}}}{\sum_{r=1}^{d} e^{v_t^{(r)}}} + \rho & if\ i = w \end{cases}$ The parameter $\rho$ captures the habit persistence and it takes values in the range [0,1] | |
| Phylum | MA(1) | AR(1) | ARMA(1,1) |
| Distinguishing characteristics | | | $\epsilon_{ijt}$ can be interpreted as arising from unobserved attributes of brands. |
| Utility framework | Yes | Yes | Yes |
| Closed-form expressions | Yes | Yes | Yes |

n.a. indicates not available.

## 4.5 Application I

**Pancake Mix Data.** Application I is based on data that was collected by a large conglomerate of noncompeting retail grocery chains located across the United States and owned by a single firm. Pancake mix purchases were recorded over a 104-week period for a subset of households that were members of these retailers' loyalty card programs. During this period, 13 unique name-brand varieties of pancake mix were available for purchase. For each customer, I observed the price paid for the purchase, the date of purchase, time of transaction, geography, unique store identification numbers, and the brand name. I included N=517 active households based on the requirement that a household has at least 30 purchases in this category during the two-year period. This resulted in a data set containing 27,148 observed choices.

**Panel Data Analysis and Time Series Analysis.** From an analytic perspective, panel data typically refers to data sets containing relatively few repeated measurements on a large number of subjects. Using **N** to indicate the number of subjects and **T** to indicate the number of repeat observations, a typical panel data set might have **N**>100 and 3<**T**<10, although these are not hard and fast rules. Alternatively, time series datasets typically contain a large number of measurements T > 50 and possibly much more on a single entity, **N**=1. Due to these differences in the form, and resulting differences in sources of variation, between and within subjects, the models used to analyze these two types of data are often distinct. When **T** is small, analysis of panels often focuses on estimating effects that are assumed common across the population while controlling for

dependence between measurements on the same subject. While the state space choice model proposed here can be parameterized as a panel model, because **T** is large I choose to implement it as a time series model, allowing all estimated effects to be separate and unique for each household in the study. This approach is consistent with the models presented in Chapter 2, and the flexibility of allowing separate parameters for each household offers an important perspective on individual-level variation in shopping patterns. Such an approach is only possible because scanner panel data typically has large numbers of repeat observations, **T**. For example, Keane (1994) analyzed scanner panels where **T** is on the order of 50 to 200 weeks. Hence, methods applied to the pancake mix data below are most appropriate to situations where analysts have both large numbers of households and large numbers of repeat observations per household.

**Models and Variable Descriptions**

Model 0: Reference Model

As a baseline for comparison, I first contemplate the model that considers utility as a function only of a brand-specific intercept and price. Specifically,

$$Y_{ijt}^{*} = \alpha_{ij} + PRICE_{ijt} \cdot \beta_i + \xi_{ijt},$$

where $\alpha_{ij}$ indicates household $i$'s intrinsic preference for brand $j$, $PRICE_{ijt}$ is the price of brand $j$ for household $i$ on occasion $t$, $\beta_i$ measures change in the marginal utility of

household $i$ with respect to price. The model is estimated using a simplified version of the procedure discussed in Chapter 2.

Model 1: State Dependence Model

To understand the importance of state dependence, I augment the reference model with a single dummy variable, $d_{ij,t-1}$, indicating the purchase of brand j in the previous time period. This is the simplest case of term 2 as described in Equation 4.2 above. I also account for serial correlation in error terms in order to avoid spurious state dependence (Erdem & Sun 2001, Heckman 1981a). Hence, Model 1 can be described with the following specification

$$Y^*_{ijt} = \alpha_{ij} + \phi_i \cdot d_{ij,t-1} + PRICE_{ijt} \cdot \beta_i + v_{ijt} + \xi_{ijt}$$

$$v_{ijt} = \lambda_i \cdot v_{ij,t-1} + \zeta_{ijt} \quad \zeta_{ijt} \sim N(0, \sigma_\zeta),$$

where $d_{ijt} = 1$ when $Y^*_{ijt} > Y^*_{ikt}\ for\ \forall\ k \neq j$, and $d_{ijt} = 0$ otherwise; $PRICE_{ijt}$ is the price paid or faced by household $i$ for brand $j$ on occasion $t$; $\alpha_{ij}$ indicates household $i$'s intrinsic preference for brand $j$; $\phi_i$ and $\lambda_i$ indicate the coefficient of the state dependence or serial correlation, respectively.

Model 2: Habit Persistence Model

In this model, the specification of the utility function differs from model 1 only in that the dummy variable for previous purchase is replaced with a lagged version of the utility. While this is not completely apparent due to the state space representation

$$Y_{ijt}^{*} = a_{ijt} + w_{ijt}$$

$$a_{ijt} = \phi_i \cdot a_{ij,t-1} + \alpha_{ij} + \beta_i \cdot x_{ijt} + (\lambda_i + \phi_i) \cdot w_{ij,t-1}$$

the equivalence was demonstrated in Section 2.2.2. As shown above, $\alpha_{ij}$ indicates household $i$'s intrinsic preference for brand $j$; $PRICE_{ijt}$ is the price paid or faced by household $i$ for brand $j$ on occasion $t$; $\alpha_{ij}$ indicates household $i$'s intrinsic preference for brand $j$; $\phi_i$ and $\lambda_i$ indicate the coefficient of the habit persistence or serial correlation, respectively.

**Data Analysis and Results**

First, I compare the performance of the three models described above and select the best fitting model for the pancake mix data. I use the Bayesian Information Criteria (BIC) to choose the best model and verify the improvement over the reference model. The BIC is widely used to compare non-nested models (Gupta & Chintagunta, 1994).

Table 4.2 reports the coefficients for state dependence, habit persistence, first order serial correlation, and price as well as log-likelihood, sample size, number of households, and the BIC statistics. In all models, I estimate the intrinsic brand-specific

effects. All of the parameter estimates in the table are weighted averages of the individual household estimates based on the means of the Bayesian posterior simulations. Weighted least squares were used to average household-level parameters in order to produce the table. I weight the individual level estimates proportional to the reciprocal of the posterior variance to control for differences in the number of purchases across households. I first consider the results for the differences in BIC (i.e., $\Delta BIC$) value between the reference model, state dependence model, and habit persistence model, respectively.

The BIC improvements for state dependence and habit persistence models are 39569.25 ($\Delta BIC = BIC_{reference} - BIC_{state\ dependence}$) and 45092.53 ($\Delta BIC = BIC_{reference} - BIC_{habit}$), respectively. Thus, I see that both model 1 and 2 show significant improvement over the reference. Furthermore, I find that the habit persistence term in model 2 gives a much better fit to the data than the state dependence term in model 1. The log-likelihood and BIC for the state dependence model are -24987.99 and 84359.62, while for the habit persistence model I obtain log likelihood and BIC values of -22226.35 and 78836.34. This makes the BIC improvement 5523.28.

As expected, all average coefficients for price are negative, indicating that consumers react negatively (switch to other brands) when prices are increased while holding all else constant.

The average coefficient for state dependence is 2.383 with standard errors of 1.186. This estimate implies that the lagged purchase has a strong effect on current decisions. The positive increment in $i$'s evaluation of the utility of purchasing brand $j$ at

occasion $t$ is $\phi_i$ if $i$ bought brand $j$ at $t-1$. If I compare two identical consumers who face the same marketing situation except that consumer A chose alternative 1 last period but consumer B did not, the current period utility evaluation of brand 1 will be roughly 2.383 units greater for consumer A than consumer B.

The coefficient for habit persistence is .275 with standard errors of .021. This points out that previous utility is positively correlated with current utility and, therefore, current choice. However, Table 4.3 shows that about 40% of consumers have negative values of habit estimates (i.e., variety-seeking behavior). I interpret this to mean that roughly 40% to 60% of households are respectively variety seeking or inertial in purchasing behaviors. In addition, the estimated coefficient of price in model 1 is -.425, while it is only -.332 (-.403) in model 2, indicating the possibility of bias in the state dependence model and considerable overestimation of the effects.

Figure 4.2 shows each individual's 95% confidence interval for phi, theta, and beta, respectively. An orange (blue) colored 95% confidence interval indicates that the interval does not contain (does contain) the value zero. In other words, the estimated value is significantly (not significantly) different from 0 at the alpha level 0.05. In comparing the intervals, I found that 107, 218, and 156 of the confidence intervals include the value 0 for phi, theta, and beta estimates, respectively. This means that about 21% to 42% of the estimated values for each parameter of habit persistence, serial correlation, and price could plausibly be 0. Due to this finding, if I fail to control for an individual's heterogeneity, I will get insignificant results from a population estimation

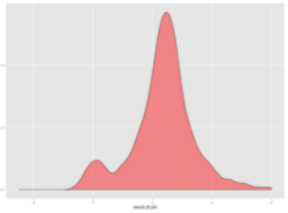
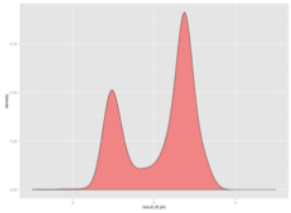
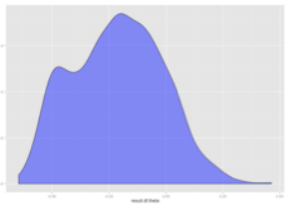
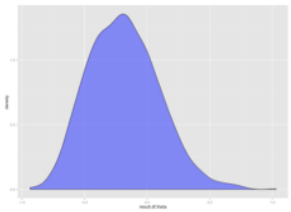
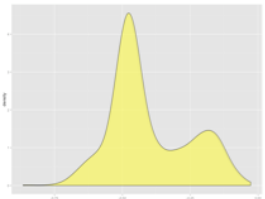
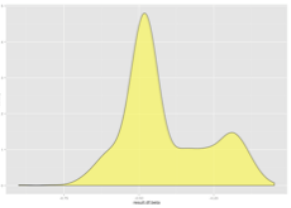
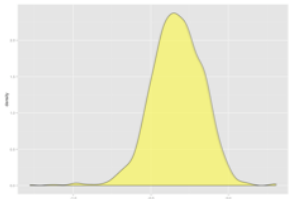
model, such as panel data analysis. As I expected, I found more insignificant confidence intervals when the estimated values are close to 0.

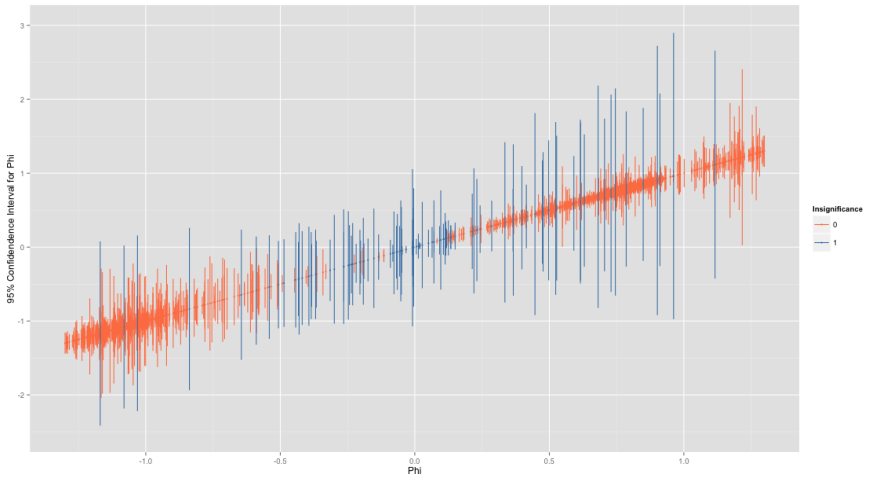
In summary, I find that the model containing habit persistence, model 2 has the largest log-likelihood and has the smallest BIC. In the context of the pancake mix data, I find that the habit persistence model is more appropriate than the state dependence model. I also find, based on model coefficients that about 40% of consumers demonstrates variety-seeking behavior. That is why, in this pancake mix category, habitual purchasing behavior (inertia) coexists with variety-seeking behavior.

**Table 4.2 - Model Comparisons: Coefficients**

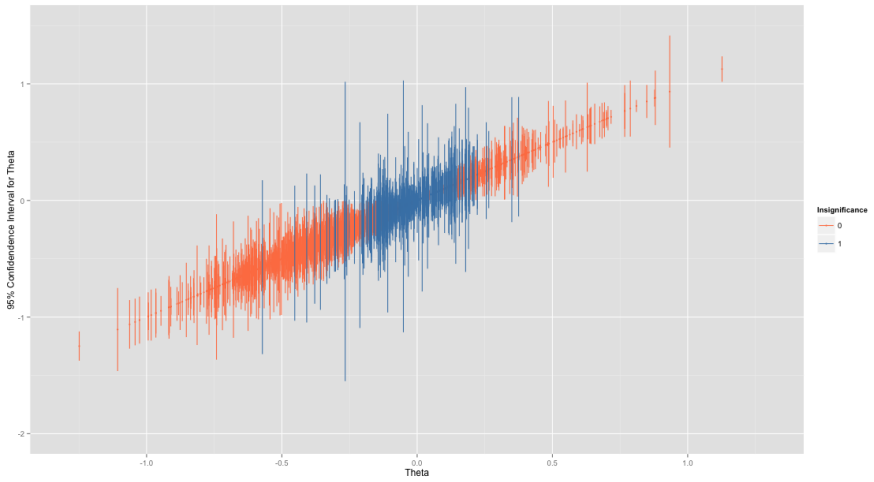
| | **Reference Model** | **State Dependence Model** | **Habit Persistence Model** |
|---|---|---|---|
| $\hat{\phi}$ | n.a. | 2.148 (.025) | .275 (.021) |
| $\hat{\lambda}$ | n.a. | -.211 (.005) | -.189 (.008) |
| $\hat{\beta}$ | -.403 (.147) | -.413 (.004) | -.319 (.005) |
| Constant | | | |
| $\alpha_1$ | .148 (.913) | .105 (.023) | .094 (.016) |
| $\alpha_2$ | .115 (1.274) | .107 (.044) | .118 (.041) |
| $\alpha_3$ | -.041 (1.891) | -.015 (.049) | .006 (.050) |
| $\alpha_4$ | .025 (.907) | .030 (.018) | .029 (.022) |
| $\alpha_5$ | .051 (.638) | .069 (.016) | .043 (.012) |
| $\alpha_6$ | .113 (.230) | .119 (.019) | .112 (.018) |
| $\alpha_7$ | .574 (.729) | .552 (.025) | .495 (.024) |
| $\alpha_8$ | .042 (.523) | .053 (.014) | .053 (.018) |
| $\alpha_9$ | -.003 (.631) | -.019 (.012) | .006 (.013) |
| $\alpha_{10}$ | .062 (.749) | .065 (.017) | .053 (.017) |
| $\alpha_{11}$ | -.0034 (.534) | .025 (.014) | .018 (.012) |
| $\alpha_{12}$ | .20 (.722) | .181 (.015) | .182 (.015) |
| Log-Likelihood | -47064.86 | -24987.99 | -22226.35 |
| BIC | 123928.87 | 84359.62 | 78836.34 |
| $\Delta BIC$ | 0 | 39569.25 | 45092.53 |
| $N$ | 517 | 517 | 517 |
| $N \times T_i$ | 27148 | 27148 | 27148 |

**Table 4.3 - Model Comparison: Distributions**

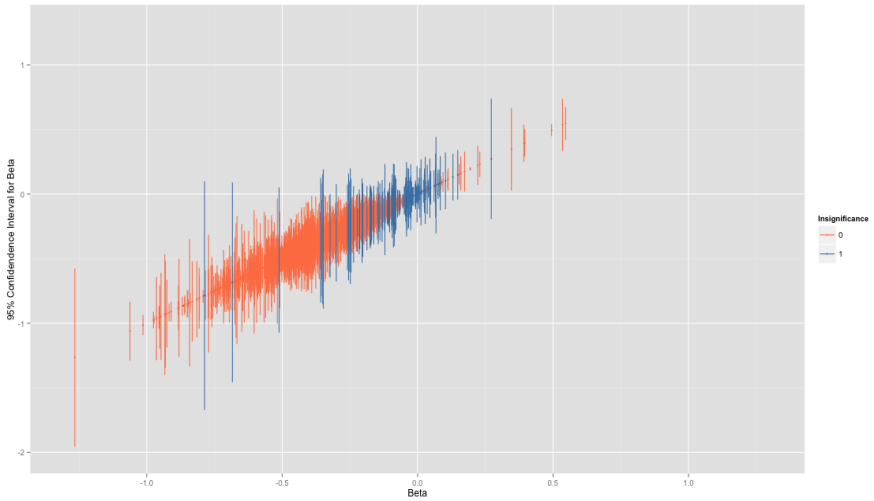
| Parameters | Null Model | State Dependence Model | Habit Persistence Model |
|---|---|---|---|
| Habit persistence term, $\hat{\phi}$. | |  |  |
| Serial correlation, $\hat{\lambda}$. | |  |  |
| Price effect, $\hat{\beta}$. |  |  |  |




(a)




(b)




(c)

**Figure 4.2 - Individual Level 95% Confidence Interval**

## 4.6 Application II

**FMCGs Data.** Fast-moving consumer goods (FMCGs) are products that are sold quickly and at relatively low cost and consist of products ranging from packaged and frozen foods a detergents to household products and other items typically found in grocery stores. The profit margin on FMCGs is relatively small, and they are generally sold in large quantities. To better understand consumer behavior, I separate FMCGs into two groups: hedonic and utilitarian goods. Choices among hedonic goods are driven by emotional desires rather than cold, cognitive deliberations. Hence, these choices represent an important domain of consumer decision making. However, much of the prior work in behavior decision theory has largely focused on the cognitive aspects of decision making without exploring its hedonic aspects (Kahneman, 1991; Khan, Dhar, & Wertenbroch, 2005). Hedonic goods are multisensory and provide for experiential consumption, fun, pleasure, and excitement (beer, for example). Utilitarian goods are primarily instrumental. Their purchase is motivated by functional aspects (e.g., cereal and soft drinks). It is important to note that both utilitarian and hedonic consumption are discretionary, and distinction between the two types of goods is a matter of degree. According to Okada (2005), hedonic consumption may be perceived as relatively more discretionary in comparison to utilitarian consumption (Khan et al., 2005). Okada finds that consumers are willing to pay more in time for hedonic goods. However, the notion of habitual purchasing is that it helps to achieve satisfaction by minimizing the costs of thinking and simplifying the decision-making process. As such, I expect that habit persistence appears to be weaker for consumers in the hedonic goods category.

**Soft Drinks.** I select 15 brands of soft drinks sold in 12 packs of 12-oz. cans. Collectively, these 15 brands account for 75 percent of market share in this category. The data covers 104 weeks and uses the entire duration for model estimation. I exclude households that do not make at least 40 transactions. A total of 436 households are included and account for 24,852 choices over two years.

**Cereal.** I select 24 brands in the 12-oz. family cereal size with category. Collectively, these 24 brands account for 81.7 percent of the cereal purchases in the category. The data covers 104 weeks, and I used an entire data set for model estimation. The same purchase criteria is applied as in the soft drink category, leading to 512 households being selected and accounting for 28,165 choices over two years.

**Beer.** I select 31 brands in the 12-oz. category. Collectively, these 24 brands account for 87 percent of the beer purchases in the category. I use the same purchase criteria for selecting families as above. A total of 381 households are selected, accounting for 16,002 choices over two years.

## Data Analysis and Results

In the interest of space, I only reported the parameter estimates of the habit persistence model and suppressed the brand-specific intercept. The parameter estimates for the habit persistence model are reported in Table 4.4.

**Soft Drinks**. The mean value of the price coefficient ($\beta$) is negative and has a correct sign. The mean value of the habit persistence parameter ($\phi$) is .576 with a standard error of .317. Table 4.5 shows that the vast majority of customers follows an

inertial purchasing pattern and with a bimodal distribution. This means that roughly 30% of households were recognized as having strong inertial purchasing behavior. There also exist strong negative serial correlations in error terms (mean of $\lambda$ = -.483 with standard error = .132).

**Cereal**. As noted earlier, I excluded the parameter of brand-specific intercepts. The mean value of the price coefficient in this model is negative (mean of $\beta$=-.176 and standard error = .200) and smaller than the mean value of the price coefficient in the soft drink data (mean of $\beta$ = -.283 and standard error= .191). The mean value of the habit persistence parameter ($\phi$) is .551, and its standard error is .312. Table 4.5 implies that a majority of consumers in this category has positive values of the habit persistence, and most of them are recognized as inertial purchasing behavior. There also exist strong and negative serial correlations in error terms (mean of $\lambda$ =-.442 with standard error = .191).

**Beer**. I reported the estimated coefficients of habit persistence, price, and serial correlation in the model. Surprisingly, Table 4.5 shows that roughly 30% of price coefficient values are negative. One explanation of this is that half of the customers in the beer category recognize the price as an indicator of quality. Additionally, about 30% of the values of the habit persistence term are negative. This means that roughly 30% of the customers in this category show variety-seeking behavior. However, the posterior mean of habit persistent term is very close to zero, .047 (.013). Weak and negative serial correlations exist in error terms (mean of $\lambda$ =-.162 with standard error = .217).

In summary, I find the existence of habitual purchasing behavior in utilitarian goods (e.g., cereal and soft drinks). Conversely, in hedonic goods (e.g., beer), I see no

evidence of habit persistence. A possible hindrance is that hedonic consumption evokes a sense of guilt (Kivetz & Simonson, 2002a, 2002b; Okada, 2005; Prelec & Loewenstein, 1998; Strahilevitz & Myers, 1998). When the sense of guilt is alleviated, hedonic consumption increases. Unlike utilitarian goods, after consumers put their effort into purchasing hedonic goods, they believe that they have earned the right to indulge in order to consume (Kivetz & Simonson, 2002a, 2002b). Guilt makes hedonic consumption more difficult to justify, and repeat purchases are less likely (Okada 2004). Several empirical studies (Verplanken, Aarts, van Knippenberg, & van Knippenberg, 1994) have shown that habits require a certain level of repetition to form and sustain them (B. Verplanken & Orbell, 2003). In the hedonic product category, I found no evidence of habit formation.

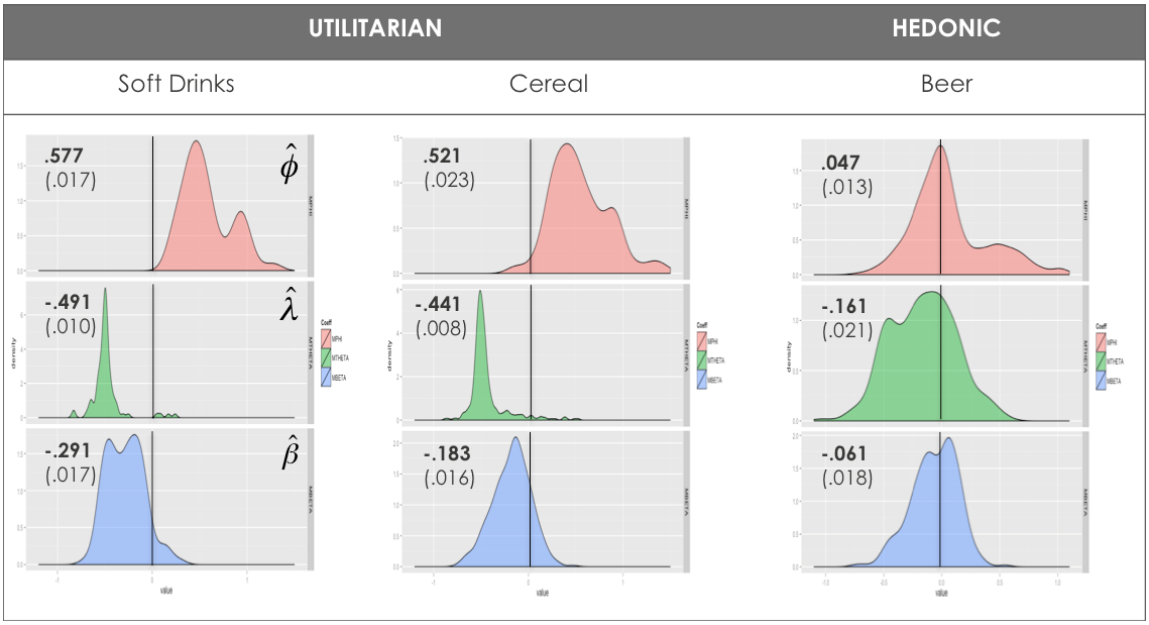
**Conclusion**

In the current chapter, I've explored the key concepts of state dependence, habit persistence, and variety seeking and their connection to customer behavior in the marketing of fast-moving consumer goods. I have also connected these concepts to their operational forms, reviewed earlier models that used these forms, and compared that earlier work to the model that I proposed in Chapter 2. Finally, I presented two case studies that indicate some of the strengths and limitations of the model while exploring actual customer behavior in scanner data.

**Table 4.4 - Utilitarian vs Hedonic: Coefficients**

| | Utilitarian | | Hedonic |
|---|---|---|---|
| | Soft drink | Cereal | Beer |
| $\hat{\phi}$ | .577 (.017) | .521 (.023) | .047 (.013) |
| $\hat{\lambda}$ | -.491 (.010) | -.441 (.008) | -.161 (.021) |
| $\hat{\beta}$ | -.291 (.017) | -.183 (.016) | -.061 (.018) |

**Table 4.5 - Utilitarian vs Hedonic: Distributions**




(a) (b) (c)

# Chapter 5

## Conclusions, Limitations, and Future Work

The goal of this dissertation was to investigate a new approach to modelling dependence across time in dynamic choice data. To accomplish this, I introduced a new state space approach to dynamic choice models and further supplied a novel fitting method. I also presented two case studies, applying the method to fast-moving consumer goods, which provided several new insights about repeat purchases in that context. I review these contributions below in brief detail before discussing limitations of the study and a variety of goals for future work in this area.

The first contribution of this dissertation was to offer an alternative to state dependence (lagged dependent variables) for capturing the phenomenon of repeat purchases observed in dynamic choice data (Roy et al. 1996). This was achieved by introducing a state space formulation of the dynamic choice model and through this model, including a lagged utility term (Heckman 1981a). The autoregressive nature of the lagged utility model provides a much richer summary of prior features and other error data as shown in Section 4.1 (Seetharaman 2004).

As I discussed briefly in Section 4.5, the model offers a great deal of flexibility. Although I have argued that, when modelling dynamic choice in the FMCG context, habit persistence through lagged utility is more appropriate than state dependence, the state space approach can easily and naturally accommodate state dependence effects with almost no additional effort. In fact, both sources of dependence could be included to

accommodate both feedback and persistence effects. As shown in Section 3.3, the model can also easily include correlated random error terms of any order in the same way it handles lagged values of utility. Beyond these effects, the state space approach can capture other forms of intervention and very general time series effects (Commandeur and Koopman, 2007). Either extreme value or normal errors can also be included in the model.

In contrast to earlier models that focused on state dependence, this approach does not suffer from an initial conditions problem, and no special effort needs to be made to deal with boundary effects of lagged variables. Because the model does not contain lagged outcome variables, it is also less sensitive to misspecification than models that contain lagged dependent variables and would not suffer from inconsistency if correlation existed in the error terms but was not accounted for. Finally, as discussed in the case studies, the model offers a natural measure of variety-seeking behavior without requiring any additional complex modelling features simply by considering the value of the state parameter $\phi$ (Seetharaman, 2004; Van Trijp, Hoyer, & Inman, 1996).

Introducing the state space model also allows for a novel fitting method, the particle filter (Doucet et al., 2001; Ridgeway & Madigan, 2003). The particle filter uses a sequential version of the importance sampling technique to integrate out unobserved states and form the conditional likelihood function for the observed data. This study is the first to use this method in a general choice modelling framework. The modelling method that I have proposed, combining state space models with particle filter fitting, exposes anyone using these techniques to an extremely broad set of structures, from

models with group effects to models that provide completely independent parameter sets for all individuals in the study.

The Bayesian approach that I used in experiments and case studies naturally allows the model to be employed for forecasting and out-of-sample studies. The sequential nature of the algorithm allows flexibility to move beyond normal error structures and explore error processes, such as *Lightning Bolt* processes (Roy et al. 1996) and more general distributions with limited modifications to the algorithm.

I applied the algorithm case studies involving sales of fast-moving consumer goods captured in scanner data furnished by a major grocery store. The studies demonstrated the wide-ranging variation in purchasing habits and price sensitivity across customers that highlights the value of the individual-level models applied here. The case studies both indicate that habit persistence is a very effective modelling variable for FMCG compared to models using only lagged variables. My second case study is also the first to use choice models to explore differences in dynamic behavior for hedonic and utilitarian goods employing choice models. I found that habit persistence was an important factor in utilitarian purchase patterns but noted an absence of habitual purchasing behavior in the hedonic category in agreement with earlier studies.
Despite the numerous potential advantages of this approach, limitations exist. First and foremost, additional testing of the models must be undertaken in both simulation and real-world situations. Previous studies of random effect choice models show that these models often produce badly biased estimates in complex situations, and both correlation, heterogeneity and choice set exclusions can play important roles (Andrews et al. 2008).

In addition, as discussed in Section 4.5, large amounts of repeat purchase data are required to fit the models that I have proposed here. In many applications, researchers will not have access to data with 30 or more choices on a single individual, as I have used here. The models considered here were coded using the scripting language *R* (R Core Team, 2012), and the method cannot currently be implemented with existing functions in any major language, to our knowledge. This approach is also computationally intensive and may require a considerable amount of run time, depending on the number of individuals in the study and the parameterization of the model.

Future research for this work falls into three categories: methodological development, marketing applications, and other applications. Within methodology, a key step is to create a more flexible model implementation that permits a number of model structures to be easily implemented. Producing a faster implementation of the software is also required in order to do extensive simulation testing. This involves development of parallel algorithms as well as coding of portions of the existing method in a compiled source, such as C++. Once a faster implementation is available, the next step is extensive simulations to understand the model performance in a wider set of situations as well as further tests of performance on real-world data. Another goal is extensions of the model to handle more complex error structures. Further extensions of the state space approach to dynamic models for count data is also appealing and could be used for modelling of store level data or total basket size in the marketing context. In all of these contexts, the development of model selection tools would be helpful.

Irrespective of the model, the Bayesian implementation allows the development of predictive models through filtering and smoothing algorithms. These predictions would allow one to estimate marginal effects of different policy and promotion changes in real applications. It would also allow the model to be useful in the context of an inventory management system.

Future work in marketing involves extending the case studies presented here to execute more thorough model comparisons, estimates of effects of price, and analyses across hedonic and utilitarian goods to better quantify factors that affect behavior. Beyond marketing, I wish to investigate inertia and evaluation mechanisms in inter-organizational partner selection. As the state space choice model can apply to organizational level constructs (Gulati and Gargiulo 1999), there is no need to find the proxy variables for organizational inertia and switching inertia.

# List of References

Akaike, Hirotugu. (1974). Markovian representation of stochastic processes and its application to the analysis of autoregressive moving average processes. *Annals of the Institute of Statistical Mathematics, 26*(1), 363-387.

Allenby, Greg M, & Lenk, Peter J. (1994). Modeling household purchase behavior with logistic normal regression. *Journal of the American Statistical Association, 89*(428), 1218-1231.

Andrews, Rick L, Ainslie, Andrew, & Currim, Imran S. (2008). On the recoverability of choice behaviors with random coefficients choice models in the context of limited data and unobserved effects. *Management Science, 54*(1), 83-99.

Ashok, Kalidas, Dillon, William R, & Yuan, Sophie. (2002). Extending discrete choice models to incorporate attitudinal and other latent variables. *Journal of Marketing Research, 39*(1), 31-46.

Athey, Susan, & Imbens, Guido W. (2007). DISCRETE CHOICE MODELS WITH MULTIPLE UNOBSERVED CHOICE CHARACTERISTICS*. *International Economic Review, 48*(4), 1159-1192.

Banerjee, Mousumi. (1994). *Influence diagnostics in longitudinal models*: University of Wisconsin--Madison.

Bates, Grace E, & Neyman, Jerzy. (1952). Contributions to the Theory of Accident Proneness. 2. True or False Contagion: DTIC Document.

Baumgartner, Bernhard. (2003). Measuring changes in brand choice behavior. *Schmalenbach Business Review, 55*, 242-256.

Bawa, Kapil. (1990). Modeling Inertia and Variety Seeking Tendencies in Brand Choice Behavior. *Marketing Science, 9*(3), 263-278.

Bucy, R, Kalman, R, & Selin, I. (1965). Comment on" The Kalman filter and nonlinear estimates of multivariate normal processes". *Automatic Control, IEEE Transactions on, 10*(1), 118-119.

Cameron, A Colin, & Trivedi, Pravin K. (2005). *Microeconometrics: methods and applications*: Cambridge university press.

Carter, Chris K, & Kohn, Robert. (1994). On Gibbs sampling for state space models. *Biometrika, 81*(3), 541-553.

Carvalho, Carlos M, Johannes, Michael S, Lopes, Hedibert F, & Polson, Nicholas G. (2010). Particle learning and smoothing. *Statistical Science, 25*(1), 88-106.

Chamberlain, G. (1978). On the Use of Panel Data. *paper presented at the Social Science Research Council Conference on life-cycle aspect of employment and the labor market, Mt. Kisco, NY*.

Chamberlain, Gary. (1984). Panel data. *Handbook of econometrics, 2*, 1247-1318.

Chandukala, Sandeep R, Kim, Jaehwan, Otter, Thomas, Rossi, Peter E, & Allenby, Greg M. (2007). Choice Models in Marketing. *Economic Assumptions, Challenges and Trends*(Now), 97-184.

Chintagunta, Pradeep, Dubé, Jean-Pierre, & Goh, Khim Yong. (2005). Beyond the endogeneity bias: The effect of unmeasured brand characteristics on household-level brand choice models. *Management Science, 51*(5), 832-849.

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

Chintagunta, Pradeep K. (1999). Variety seeking, purchase timing, and the "lightning bolt" brand choice model. *Management Science, 45*(4), 486-498.

Chintagunta, Pradeep K. (2001). Endogeneity and heterogeneity in a probit demand model: Estimation using aggregate data. *Marketing Science, 20*(4), 442-456.

Coleman, James S. (1964). *Models of change and response uncertainty*: Prentice-Hall Englewood Cliffs, NJ.

Commandeur, Jacques JF, & Koopman, Siem Jan. (2007). *An introduction to state space time series analysis*: Oxford University Press.

Commandeur, Jacques JF, Koopman, Siem Jan, & Ooms, Marius. (2011). Statistical software for state space methods. *Journal of statistical software, 41*(1), 1-18.

Constantinides, George M. (1990). Habit formation: A resolution of the equity premium puzzle. *Journal of political Economy*, 519-543.

Corstjens, Marcel, & Lal, Rajiv. (2000). Building store loyalty through store brands. *Journal of Marketing Research, 37*(3), 281-291.

Doucet, Arnaud, De Freitas, Nando, & Gordon, Neil. (2001). *Sequential Monte Carlo methods in practice*: Springer.

Doucet, Arnaud, Godsill, Simon, & Andrieu, Christophe. (2000). On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and computing, 10*(3), 197-208.

Durbin, James, & Koopman, Siem Jan. (2000). Time series analysis of non□Gaussian observations based on state space models from both classical and Bayesian perspectives. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 62*(1), 3-56.

Durbin, James, & Koopman, Siem Jan. (2012). *Time series analysis by state space methods*: Oxford University Press.

Erdem, Tülin, & Keane, Michael P. (1996). Decision-making under uncertainty: Capturing dynamic brand choice processes in turbulent consumer goods markets. *Marketing Science, 15*(1), 1-20.

Erdem, Tülin, & Sun, Baohong. (2001). Testing for choice dynamics in panel data. *Journal of Business & Economic Statistics, 19*(2), 142-152.

Erdem, Tülin, & Winer, Russell S. (1998). Econometric modeling of competition: A multi-category choice-based mapping approach. *Journal of Econometrics, 89*(1), 159-175.

Fishman, G. (2005). A First Course in Monte Carlo. *Thomson Brooks/Cole*.

Frühwirth-Schnatter, Sylvia, & Frühwirth, Rudolf. (2007). Auxiliary mixture sampling with applications to logistic models. *Computational Statistics & Data Analysis, 51*(7), 3509-3528.

Gelman, Andrew, Carlin, John B, Stern, Hal S, Dunson, David B, Vehtari, Aki, & Rubin, Donald B. (2013). *Bayesian data analysis*: CRC press.

Gerlach, Richard, Carter, Chris, & Kohn, Robert. (2000). Efficient Bayesian inference for dynamic mixture models. *Journal of the American Statistical Association, 95*(451), 819-828.

Gordon, Neil J, Salmond, David J, & Smith, Adrian FM. (1993). *Novel approach to nonlinear/non-Gaussian Bayesian state estimation.* Paper presented at the IEE Proceedings F (Radar and Signal Processing).

Greene, William H. (2003). Econometric analysis, 5th. *Ed.. Upper Saddle River, NJ*.

Guadagni, Peter M, & Little, John DC. (1983). A logit model of brand choice calibrated on scanner data. *Marketing Science, 2*(3), 203-238.

Gulati, Ranjay, & Gargiulo, Martin. (1999). Where do interorganizational networks come from? 1. *American journal of sociology, 104*(5), 1439-1493.

Gupta, Sachin, & Chintagunta, Pradeep K. (1994). On using demographic variables to determine segment membership in logit mixture models. *Journal of Marketing Research*, 128-136.

Haaijer, Rinus, & Wedel, Michel. (2001). Habit persistence in time series models of discrete choice. *Marketing Letters, 12*(1), 25-35.

Harvey, Andrew C. (1990). *Forecasting, structural time series models and the Kalman filter*: Cambridge university press.

Hausman, Jerry, & McFadden, Daniel. (1984). Specification tests for the multinomial logit model. *Econometrica: Journal of the Econometric Society*, 1219-1240.

Heckman, James J. (1977). Dummy endogenous variables in a simultaneous equation system: National Bureau of Economic Research Cambridge, Mass., USA.

Heckman, James J. (1981a). Heterogeneity and state dependence *Studies in labor markets* (pp. 91-140): University of Chicago Press.

Heckman, James J. (1981b). Statisticalmodelsfordiscretepaneldata.C.F. Manski, D. McFadden, eds. Structural Analysis of Discrete Data with Applications. .*MIT Press, Cambridge, MA*, 179-195.

Heckman, James J, & Willis, Robert J. (1977). A beta-logistic model for the analysis of sequential labor force participation by married women. *The Journal of Political Economy*, 27-58.

Heiss, Florian. (2008). Sequential numerical integration in nonlinear state space models for microeconometric panel data. *Journal of Applied Econometrics, 23*(3), 373-389.

Heiss, Florian, & Winschel, Viktor. (2008). Likelihood approximation by numerical integration on sparse grids. *Journal of econometrics, 144*(1), 62-80.

Hensher, David A, Rose, John M, & Greene, William H. (2005). *Applied choice analysis: a primer*: Cambridge University Press.

Honoré, Bo E, & Tamer, Elie. (2006). Bounds on parameters in panel dynamic discrete choice models. *Econometrica, 74*(3), 611-629.

Hsiao, Cheng. (2003). *Analysis of panel data* (Vol. 34): Cambridge university press.

Hsiao, Cheng, Hashem Pesaran, M, & Kamil Tahmiscioglu, A. (2002). Maximum likelihood estimation of fixed effects dynamic panel data models covering short time periods. *Journal of econometrics, 109*(1), 107-150.

Jacoby, Jacob, & Chestnut, Robert W. (1978). Brand loyalty measurement and management.

Kahneman, Daniel. (1991). Commentary: judgment and decision making: a personal view. *Psychological science*, 142-145.

Kailath, Thomas, Sayed, Ali H, & Hassibi, Babak. (2000). Linear estimation.

Kalman, Ra E. (1962). *Control of randomly varying linear dynamical systems.* Paper presented at the Proceedings of symposia in applied mathematics.

Kalman, Rudolf E. (1962). Canonical structure of linear dynamical systems. *Proceedings of the National Academy of Sciences of the United States of America, 48*(4), 596.

Kanetkar, Vinay, Weinberg, Charles B., & Weiss, Doyle L. (1990). An Empirical Analysis of Alternative Definitions of Brand Loyalty in Choice Models. *Discussion Paper for Banff Invitational Symposium on Consumer Decision Making and Choice Behavior.*

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

Keane, Michael P. (2013). Panel Data Discrete Choice Models of Consumer Demand. *Prepared for The Oxford Handbooks: Panel Data*.

Khan, Uzma, Dhar, Ravi, & Wertenbroch, Klaus. (2005). A behavioral decision theory perspective on hedonic and utilitarian choice. *Inside consumption: Frontiers of research on consumer motives, goals, and desires*, 144-165.

Kitamura, Ryuichi. (1990). Panel analysis in transportation planning: An overview. *Transportation Research Part A: General, 24*(6), 401-415.

Kitamura, Ryuichi, & Bunch, David S. (1990). Heterogeneity and state dependence in household car ownership: A panel analysis using ordered-response probit models with error components. *University of California Transportation Center*.

Kivetz, Ran, & Simonson, Itamar. (2002a). Earning the right to indulge: effort as a determinant of customer preferences toward frequency program rewards. *Journal of Marketing Research, 39*(2), 155-170.

Kivetz, Ran, & Simonson, Itamar. (2002b). Self□Control for the Righteous: Toward a Theory of Precommitment to Indulgence. *Journal of Consumer Research, 29*(2), 199-217.

Layton, L. (1978). Unemployment Over the Work History. *Ph.D. dissertation, Department of Economics, Columbia University, New York*.

Lopes, Hedibert F, & Tsay, Ruey S. (2011). Particle filters and Bayesian inference in financial econometrics. *Journal of Forecasting, 30*(1), 168-209.

Marin, Jean-Michel, & Robert, Christian. (2007). *Bayesian Core: A Practical Approach to Computational Bayesian Statistics: A Practical Approach to Computational Bayesian Statistics*: Springer.

McAlister, Leigh, Srivastava, Rajendra, Horowitz, Joel, Jones, Morgan, Kamakura, Wagner, Kulchitsky, Jack, . . . Yai, Tetsuo. (1991). Incorporating choice dynamics in models of consumer behavior. *Marketing Letters, 2*(3), 241-252.

McCormick, Tyler H, Raftery, Adrian E, Madigan, David, & Burd, Randall S. (2012). Dynamic logistic regression and dynamic model averaging for binary classification. *Biometrics, 68*(1), 23-30.

McFadden, Daniel. (1974). The measurement of urban travel demand. *Journal of public economics, 3*(4), 303-328.

Mellens, M. (1996). in Marketing.

Miranda, Alfonso. (2007). *Dynamic Probit models for panel data: A comparison of three methods of estimation.* Paper presented at the United Kingdom Stata Users' Group Meetings.

Okada, Erica Mina. (2005). Justification effects on consumer choice of hedonic and utilitarian goods. *Journal of Marketing Research, 42*(1), 43-53.

Petris, Giovanni, Campagnoli, Patrizia, & Petrone, Sonia. (2009). *Dynamic linear models with R*: Springer.

Prelec, Drazen, & Loewenstein, George. (1998). The red and the black: Mental accounting of savings and debt. *Marketing Science, 17*(1), 4-28.

Ridgeway, Greg, & Madigan, David. (2003). A sequential Monte Carlo method for Bayesian analysis of massive datasets. *Data Mining and Knowledge Discovery, 7*(3), 301-319.

Rizzo, Maria L. (2007). Statistical Computing with R. *Chapman & Hall/CRC* (The R Series).

Roberts, Gareth O, Gelman, Andrew, & Gilks, Walter R. (1997). Weak convergence and optimal scaling of random walk Metropolis algorithms. *The annals of applied probability, 7*(1), 110-120.

Roy, Rishin, Chintagunta, Pradeep K, & Haldar, Sudeep. (1996). A framework for investigating habits,"The Hand of the Past," and heterogeneity in dynamic brand choice. *Marketing Science, 15*(3), 280-299.

Seetharaman, PB. (2004). Modeling multiple sources of state dependence in random utility models: A distributed lag approach. *Marketing Science, 23*(2), 263-271.

Seetharaman, PB, & Chintagunta, Pradeep. (1998). A model of inertia and variety-seeking with marketing variables. *International Journal of Research in Marketing, 15*(1), 1-17.

Shalizi, Cosma Rohilla. (Forthcoming). Advanced Data Analysis from an Elementary Point of View.

Strahilevitz, Michal, & Myers, John G. (1998). Donations to charity as purchase incentives: How well they work may depend on what you are trying to sell. *Journal of Consumer Research, 24*(4), 434-446.

Train, Kenneth E. (2009). *Discrete choice methods with simulation*: Cambridge university press.

Tsay, Ruey S. (2005). *Analysis of financial time series* (Vol. 543): John Wiley & Sons.

Van Trijp, Hans CM, Hoyer, Wayne D, & Inman, J Jeffrey. (1996). Why switch? Product category: level explanations for true variety-seeking behavior. *Journal of Marketing Research*, 281-292.

Venkatesan, Rajkumar, & Farris, Paul W. (2012). Measuring and managing returns from retailer-customized coupon campaigns. *Journal of marketing, 76*(1), 76-94.

Verplanken, Aarts, H, van Knippenberg, A, & van Knippenberg, C. (1994). Attitude versus general habit. *Journal of Applied Social Psychology, 24*, 285-300.

Verplanken, Bas, & Orbell, Sheina. (2003). Reflections on Past Behavior: A Self□Report Index of Habit Strength1. *Journal of Applied Social Psychology, 33*(6), 1313-1330.

West, M, & Harrison, P.J. (1997). Bayesian Forecasting and Dynamic Models. *Springer, 2nd*.

West, Mike. (1987). On scale mixtures of normal distributions. *Biometrika, 74*(3), 646-648.

West, Mike, Harrison, P Jeff, & Migon, Helio S. (1985). Dynamic generalized linear models and Bayesian forecasting. *Journal of the American Statistical Association, 80*(389), 73-83.

Wooldridge, Jeffrey M. (2005). Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity. *Journal of applied econometrics, 20*(1), 39-54.

Wooldridge, Jeffrey M. (2010). *Econometric analysis of cross section and panel data*: MIT press.

# Vita

Kang Bok Lee was born in Seoul, South Korea on March 18, 1980, to the parents Hoonkoo Lee and Myunghee Choi. He has an older sister, Sora Lee. Kang married Sumin Han on July 7th 2007. Kang obtained a Bachelor of Science degree in Mathematics and Economics from SungKyunKwan University in August of 2005, and a Master of Statistics degree from University of California at Riverside in August of 2009.

In August of 2010, Kang entered the PhD program at the University of Tennessee. He has accepted a tenure-track assistant professor faculty position at the Raymond J. Harbert College of Business at Auburn University in Auburn, Alabama and will move there after his August, 2014 graduation from the University of Tennessee.