DATA AUGMENTATION FOR LATENT VARIABLES IN MARKETING

DISSERTATION

Presented in Partial Fulfillment of Requirements for
the Degree Doctor of Philosophy in the Graduate
School of The Ohio State University

By
Ling-Jing Kao, B.A., M.S.

* * * * *

The Ohio State University
2006

Dissertation Committee:
Dr. Greg M. Allenby, Adviser
Dr. H. Rao Unnava
Dr. Thomas Otter

Approved by
Adviser
Business Administration Graduate Program
ABSTRACT

Latent variable models are an important aspect of consumer research whenever the determinants of behavior are an important aspect of study. The purpose of this thesis is to develop a new method of error augmentation to deal with latent variable models of heterogeneous, non-linear consumer behavior for two issues commonly encountered in marketing. The first issue relates to a consumer’s purchase decision driven by a time-varying latent behavior process. The second issue relates to consumer preferences affected by multiple unobserved factors.

This thesis comprises three essays. The first issue is addressed in the first and the second essays, and the second issue is addressed in the third essay. The new method of error augmentation is applied to estimate models proposed in this thesis. The new method of error augmentation is needed because, in the proposed models, the observed discrete choices do not have a direct correspondence to the errors. The proposed models would be difficult to estimate without the new approach.

The first essay develops a new method of error augmentation for state-space models of economic behavior where the observed behavior is related to a latent variable whose temporal variation is described by a state equation. The proposed state-space model is applied to analyze a consumer’s purchase and resignation decisions in a membership club.
The result indicates that increasing inter-arrival time between shipments can lead to longer customer longevity and greater sales.

The second essay investigates an alternative method of modeling customer inter-purchase times. A state-space model is proposed to investigate the possibility to model inter-purchase times as an independent variable. The results indicate that the proposed state-space model can accurately describe customer behavior when the specification of the state equation is plausible for the data.

In the third essay, a demand model is developed to address three issues in choice modeling. The first issue relates to the effects of multiple treatments for data collected in a pre-post study. The second issue relates to a marketing action of line extension that is widely adopted in marketing practice. The last issue relates to consumer decisions of brand-pack and no-choice for consumer packaged goods at the level of stock-keeping unit.

Data from a leading consumer packaged goods company are used to study changes in consumer preferences and sensitivities in a simulated shopping environment. The results indicate that consumers’ reactions to media are very heterogeneous. Media can make some consumers have extreme preferences, and make preferences of some consumers become more homogeneous.

This thesis contributes marketing literature by developing a new method of error augmentation for latent variable models that cannot be estimated by standard approaches. The new method of error augmentation is illustrated by three different marketing applications in this thesis. The state-space model proposed in the first and the second essays can be extended to study consumer learning or consumer searching behavior. The
demand model proposed in the third essay can be extended to study consumer preference changes in multiple stages.
Dedicated to my parents Tsung-Ching Kao and Yu-Hsiu Hsu
ACKNOWLEDGMENTS

I wish to express sincere thanks to my advisor, Dr. Greg M. Allenby, for the time and effort he put into my doctoral education. You are an outstanding scholar. You have been a responsible facilitator, gatekeeper, and protector in my graduate study as well as a wonderful friend in my life. I have learned tremendously from you.

I also want to thank you for the challenges and frustrations you give to me in research. These challenges and frustrations make me think a lot about my life and myself. It makes me be tougher and stronger in the road of pursuing my dream. Without these challenges and frustrations, I will still be a child spoiled by people around me. This training process of doctoral education has installed me a dedication to rigor in research. Without your mentoring, I could not have succeeded in my doctoral journey.

My appreciation is also extended to the other members of my dissertation committee, Dr. H. Rao Unnava and Dr. Thomas Otter. I want to thank them for providing guidance and support during my dissertation research and during my time at Ohio State. I thank the other marketing faculty at Ohio State for their enduring support to my doctoral education. I want to thank current and past Ph.D. students particularly Jaehwan Kim, Yancy Edwards, Tim Gilbrid, Sandeep Chandukala, and Jeff Dotson for their support and friendship.
I also want to thank Cindy Coykendale and Lisa Gang for providing invaluable help on all administrative details. I want to thank Tim Renken and June Hahn for providing data for my dissertation. My dissertation cannot be completed on time without the help from Curtis Smith in the department of computing and communication services. I want to thank you for setting up R environment in Unix servers for me.

I could not have started or completed my doctoral studied without the support of my family. My parents Tsung-Ching Kao and Yu-Hsiu Hsu always stand by for me with strong faith while I pursued my dreams and for being patient with, believing in, and walking with me. I also appreciate my brothers and sister, Yu-Sui, Kuo-Ting, and Hsin-Chih, for their overwhelming concern and encouragement.

Finally, I would like to give my special thank to Dr. Chih-Chou Chiu in National Taipei University of Technology for his invaluable friendship and encouragement along the way. I thank you to stand by for me and listen to me while I was in depression. You have tremendous influence on my decision of pursuing doctoral degree. Your humanity and personality have inspired me to contribute myself to our society and people in the world.
VITA

November 29, 1974...............................Born – Taipei, Taiwan

1997......................................................B. A., Business Administration
                                Fu-Jen Catholic University, Taipei, Taiwan

2001......................................................M. S., Statistics
                                Texas A&M University, College Station, TX, USA

2001-present........................................Graduate Teaching and Research Associate,
                                The Ohio State University

FIELDS OF STUDY

Major Field: Business Administration
Specialization: Marketing
TABLE OF CONTENTS

Abstract .......................................................................................................................... ii
Dedication ...................................................................................................................... v
Acknowledgements ........................................................................................................ vi
Vita................................................................................................................................. viii
List of Tables ................................................................................................................. xi
List of Figures ................................................................................................................. xii

Chapters:

1. Introduction................................................................................................................ 1

2. Data augmentation and latent variable models ......................................................... 6

   3.1 Introduction..................................................................................................... 12
   3.2 State-space models for economic behavior..................................................... 14
       3.2.1 Model estimation ................................................................................ 17
       3.2.2 Model identification............................................................................ 24
       3.3.3 Simulation study ................................................................................. 26
   3.3 Direct marketing application........................................................................... 27
   3.4 Estimation results............................................................................................ 30
   3.5 Discussion....................................................................................................... 32
   3.6 Conclusion remarks ........................................................................................ 33

   4.1 Introduction..................................................................................................... 46
   4.2 Model development ........................................................................................ 48
   4.3 Data and model specification.......................................................................... 51
       4.3.1 State-space model specification.......................................................... 52
       4.3.2 Inter-purchase time model specification............................................. 53
   4.4 Parameter estimates and predictive results ..................................................... 54
   4.5 Discussion....................................................................................................... 56

5. Essay 3: Modeling Media Interactions and Preference Change in Panel Data........ 64
   5.1 Introduction..................................................................................................... 64
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.2</td>
<td>Literature review</td>
<td>70</td>
</tr>
<tr>
<td>5.2.1</td>
<td>Preference change and consumer heterogeneity</td>
<td>70</td>
</tr>
<tr>
<td>5.2.2</td>
<td>Advertising and media effects</td>
<td>72</td>
</tr>
<tr>
<td>5.2.3</td>
<td>Discrete quantity</td>
<td>75</td>
</tr>
<tr>
<td>5.3</td>
<td>Model development</td>
<td>80</td>
</tr>
<tr>
<td>5.3.1</td>
<td>Treatment effect</td>
<td>84</td>
</tr>
<tr>
<td>5.3.2</td>
<td>Generating latent utility</td>
<td>88</td>
</tr>
<tr>
<td>5.3.3</td>
<td>Data augmentation for error terms</td>
<td>91</td>
</tr>
<tr>
<td>5.4</td>
<td>Empirical application</td>
<td>92</td>
</tr>
<tr>
<td>5.4.1</td>
<td>Data of consumer packaged goods</td>
<td>92</td>
</tr>
<tr>
<td>5.4.2</td>
<td>Proposed models for the empirical study</td>
<td>95</td>
</tr>
<tr>
<td>5.4.3</td>
<td>Location identification of proposed models</td>
<td>98</td>
</tr>
<tr>
<td>5.5</td>
<td>Results</td>
<td>100</td>
</tr>
<tr>
<td>5.5.1</td>
<td>Model comparison</td>
<td>100</td>
</tr>
<tr>
<td>5.5.2</td>
<td>Coefficient estimates</td>
<td>102</td>
</tr>
<tr>
<td>5.6</td>
<td>Conclusions</td>
<td>109</td>
</tr>
<tr>
<td>6</td>
<td>Conclusions</td>
<td>152</td>
</tr>
<tr>
<td>Appendices</td>
<td></td>
<td>156</td>
</tr>
<tr>
<td>Appendix A</td>
<td>MCMC Estimation for Essay1</td>
<td>156</td>
</tr>
<tr>
<td>Appendix B</td>
<td>MCMC Estimation for Essay 2</td>
<td>166</td>
</tr>
<tr>
<td>Appendix C</td>
<td>MCMC Estimation for Essay 3</td>
<td>175</td>
</tr>
<tr>
<td>List of references</td>
<td></td>
<td>215</td>
</tr>
</tbody>
</table>
## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1 Parameter estimates</td>
<td>45</td>
</tr>
<tr>
<td>4.1 Parameter estimates (posterior standard deviations)</td>
<td>63</td>
</tr>
<tr>
<td>5.1 Levels of independent variables</td>
<td>134</td>
</tr>
<tr>
<td>5.2 Descriptive statistics</td>
<td>135</td>
</tr>
<tr>
<td>5.3 Brand switching matrices</td>
<td>136</td>
</tr>
<tr>
<td>5.4 The frequency of media exposures</td>
<td>138</td>
</tr>
<tr>
<td>5.5 Sticker</td>
<td>139</td>
</tr>
<tr>
<td>5.6 Number of respondents who do not select media of each brand</td>
<td>140</td>
</tr>
<tr>
<td>5.7 Model comparison</td>
<td>141</td>
</tr>
<tr>
<td>5.8 Posterior estimates of $\beta$</td>
<td>142</td>
</tr>
<tr>
<td>5.9 Posterior estimates of $V_\beta$</td>
<td>143</td>
</tr>
<tr>
<td>5.10 Posterior estimates of $\gamma$</td>
<td>144</td>
</tr>
<tr>
<td>5.11 Posterior estimates of $V_\gamma$</td>
<td>146</td>
</tr>
<tr>
<td>5.12 Posterior estimates of $\theta$</td>
<td>148</td>
</tr>
<tr>
<td>5.13 Posterior estimates of $V_\theta$</td>
<td>150</td>
</tr>
</tbody>
</table>
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Markov chain realizations of $\alpha$: (a) New algorithm; (b) Standard algorithm</td>
</tr>
<tr>
<td>3.2</td>
<td>Identification analysis for latent inventory ($s$) and effect size ($\beta$)</td>
</tr>
<tr>
<td>3.3</td>
<td>Markov chain realizations of model parameters for simulation study</td>
</tr>
<tr>
<td>3.4</td>
<td>Descriptive statistics</td>
</tr>
<tr>
<td>3.5</td>
<td>Posterior distribution of customer and item effects</td>
</tr>
<tr>
<td>3.6</td>
<td>Acceptance rates versus items effects ($\alpha_k$)</td>
</tr>
<tr>
<td>3.7</td>
<td>Posterior distribution of autocorrelation coefficients ($\phi$)</td>
</tr>
<tr>
<td>3.8</td>
<td>Customer longevity ($T_j$) versus autocorrelation coefficients ($\phi$)</td>
</tr>
<tr>
<td>3.9</td>
<td>Posterior distribution of initial state ($s_0$)</td>
</tr>
<tr>
<td>3.10</td>
<td>Expected demand for offering inter-arrival times</td>
</tr>
<tr>
<td>4.1</td>
<td>Comparison between standard inter-purchase time model and state-space model</td>
</tr>
<tr>
<td>4.2</td>
<td>Heterogeneity distribution of state-space model parameters</td>
</tr>
<tr>
<td>4.3</td>
<td>Heterogeneity distribution of inter-purchase time model parameters</td>
</tr>
<tr>
<td>4.4</td>
<td>Comparison of model forecasts</td>
</tr>
<tr>
<td>5.1</td>
<td>Box plots for the posterior mean of brand preference of multiplicative Model</td>
</tr>
<tr>
<td>5.2</td>
<td>Scatter plots for pre-post posterior mean of brand intercept of multiplicative model</td>
</tr>
</tbody>
</table>
5.3 Scatter plots for posttest brand intercept of Brand A+ and pretest brand preference intercept of established brands ................................................................. 113

5.4 Scatter plots for posttest brand intercept of Brand A+ and posttest brand intercept of established brands .............................................................................. 114

5.5 Components of consumer preferences intercept of Brand A+ ($\beta_{01,h}$) .......... 115

5.6 Histogram for the difference of pre-post posterior mean of consumer sensitivities to marketing merchandising variables of multiplicative model....... 116

5.7 Histogram for the difference of pre-post posterior mean of consumer preferences to product attributes 1,2 and 3 of multiplicative model ...................... 117

5.8 Histogram for the difference of pre-post posterior mean of consumer preferences to product attributes 4 of multiplicative model ............................... 118

5.9 Histogram for the difference of pre-post posterior mean of consumer preference to quantity ($\beta_{x,h}$) and the outside goods ($\beta^*_{T,h}$) of multiplicative model ......... 119

5.10 Box plots for the posterior mean of $\gamma_h$ for all information sources of Brand A+. 120

5.11 Media effects on the intercept of Brand A .................................................................................................................. 121

5.12 Media effects on the intercept of Brand B ................................................................................................................. 123

5.13 Media effects on the intercept of Brand D .................................................................................................................. 125

5.14 Aggregate media effects ($\theta_{z,h} M_h$) ................................................................................................................. 127

5.15 Media effects on consumer preferences to product attribute 1 ...................... 128

5.16 Media effects on consumer preferences to $\ln(x+1)$ ........................................ 130

5.17 Media effects on consumer sensitivities to $\ln(T-p(x))$ ................................ 132
CHAPTER 1

INTRODUCTION

Marketing data reflect latent behavioral processes that are heterogeneous and non-linear. Consumers are often thought to make choices according to principles of utility maximization once they attend to marketing stimuli. However, factors that influence behavior are often not observed by the researcher, and the factors that influence the engagement of attention are not well represented by a linear compensatory model. For example, household inventories non-linearly affect brand preference and purchase timing in the presence of diminishing marginal returns. When inventories are not observed, complications arise in estimating demand models because the data are serially dependent unless restrictive assumptions are made about specific inventory levels at each point in time. Likewise, consumer preferences for goods can exhibit temporal changes when inventions such as learning take place. Complications arise in tracking latent preference changes at the individual-level because of the relatively short panel lengths present in marketing application.

The purpose of this thesis is to develop methods of dealing with heterogeneous, non-linear models of behavior for problems commonly encountered in marketing. The dissertation comprises two major themes, one focused on a model for decision making (i.e., the likelihood) and the other dealing with the distribution of heterogeneity. The
first theme focuses on state-space models of economic behavior where a latent state variable stochastically evolves over time and is an argument of a household's utility function. Observed choices are assumed to be related to marginal utility, giving rise to a class of models where the state and observation equations share common parameters and error realizations. The second theme concerns random-effects specifications of heterogeneity where pre-post measurements are available to the researcher. Since the pre-post measurements are a pair of observations collected from a respondent, the treatment effect is studied by relating pre-post measurements to the same random-effect realization.

Data augmentation was originally introduced into the statistics literature by Tanner and Wong (1987) as a method of simplifying computations associated with properties (e.g., moments) of the posterior distribution. Albert and Chib (1993) developed the application of data augmentation to estimate the probit model where the observed data are viewed as censored realizations of latent utility. Augmentation methods are used to simplify analysis in hierarchical Bayes models, where the augmented variables are treated as unobserved parameters. According to Bayesian theorem, researchers compute the joint posterior distribution of the augmented variables and other parameters, and then margin down to the posterior distribution of parameters of interest.

Error augmentation—a new variant of data augmentation—and a new estimation algorithm are developed in this thesis to estimate latent variable models proposed in this thesis. The standard data augmentation cannot be applied to the proposed models because the observed discrete choices do not have direct correspondence to the errors. As a result, the errors cannot be generated directly form a distribution, and the likelihood functions of
proposed models are difficult to compute without the new approach. The discussion of data augmentation and the proposed method are provided in next chapter.

The thesis comprises three essays. The first essay develops a new method of data augmentation for state-space models for economic behavior where the observed behavior is related to a latent variable whose temporal variation is described by a state equation. This new method is needed for models of economic behavior because the observed data are assumed to be associated by marginal utilities that are influenced by the state variables, leading to shared parameters and error realizations in the observation and state equations. The new algorithm simulates realizations of the state variable according to Bayes rule, and then uses the realizations to construct corresponding realizations of error terms. These error terms are then used to reconstruct state variables for different values of parameters. This simulation procedure simplifies the high-dimensional analysis associated with the estimation of latent variable models.

Properties of the proposed estimator are demonstrated in two simulation studies. The studies show that the proposed method can deal with complicated model structures that cannot be estimated with standard methods. Direct marketing data from a membership program are used to illustrate the method, where two observation equations are used to represent the purchase and resignation decisions of customers, and a state equation is used to represent stochastic variation of latent inventory. The result indicates that the proposed method provides a flexible framework for analyzing economic models of behavior in marketing.

The second essay investigates an alternative method of modeling customer inter-purchase times. In traditional models of direct marketing, inter-purchase times are treated
as dependent variables whose model parameters are used to identify profitable customers. In this essay, models that treat purchase timing as an independent variable are explored. A latent inventory model is developed according to the assumption that purchases are triggered by inventories below a threshold value. The specification of this model is different from the model for the membership data in the first essay, and is explored using two direct marketing datasets. The first dataset is from an office supply company engaged in business-to-business selling in the United States. The second dataset is from a direct marketing company specializing in cosmetics, shampoo, toothpaste and food supplements selling in Taiwan. The performance of the proposed model is compared to a traditional inter-purchase time model, with results supporting the proposed model in the business-to-customer dataset which comprises more regular behavior of customers.

The third essay develops a model with random-effect specification of heterogeneity for a pretest-posttest study. The measurement of a dependent variable are collected twice from a respondent. In traditional pre-post measurements, treatment effects are evaluated by subtracting the post measurement from the pre measurement to remove subject-specific effects. Pre-post measurements within a random-effects model are achieved by relating both measurements to the same random-effect realizations. The new method of error augmentation developed in this thesis is needed to implement the model with no-choice decisions at the level of stock keeping unit. Since no-choice decisions lead to partial ranks among utilities of available items, there is no direct correspondence between the observed choices and the errors. The likelihood cannot be evaluated by the standard approach. Data from a leading packaged goods company are used to illustrate the method by investigating changes in consumer preference and sensitivities in a
simulated shopping environment. The purpose of this study is to explore the effect of brand extension, the impact of media on the likelihood of purchasing a new brand, and changes in consumer preferences and sensitivities to marketing stimuli.

The reminder of this thesis is organized as follows. In Chapter 2, the literature of data augmentation and choice model with latent variables is discussed, and a new variant of data augmentation is introduced. In Chapter 3, the first essay “State-Space Model for Economics Behavior” is included. The second essay “A State-Space Model of Purchase Timing for Direct Marketing” is presented in Chapter 4. Chapter 5 presents the third essay “Modeling Media Interactions and Preference Change in the Panel Data”. Chapter 6 offers a discussion and contribution of this thesis to the literature.
CHAPTER 2

DATA AUGMENTATION AND LATENT VARIABLE MODELS

The method of data augmentation is originally proposed by Tanner and Wong (1987). It provides a scheme to augment the observed data $y$ by latent variable $z$. For example, a model is specified as $y = f(\theta)$ in which the posterior distribution $p(\theta | y)$ is difficult to estimate directly. The method of data augmentation suggests introducing a latent variable $z$ to estimate $p(\theta | y, z)$. By integrating out $z$ from $p(\theta | y, z)$, the posterior distribution $p(\theta | y)$ can be obtained. The implementation of data augmentation method is straightforward in a Bayesian framework since Bayesian views all the unknown variables as parameters. The estimation can be processed by drawing $z$ and $\theta$ from their conditional distributions $p(z | y, \theta)$ and $p(\theta | y, z)$ iteratively.

Consider the example of binary choice in which the binary choice $y_i$ is observed. $y_i$ equals to 1 when the purchase is observed. Otherwise, $y_i$ equals to 0. Consumers are assumed to be utility optimizer. If the marginal utility $z_i$ is above a threshold, a consumer will purchase. Otherwise, a consumer will not purchase. The marginal utility $z_i$ is a function of product attributes, marketing activities, and error terms which capture the effect of other unobserved factors. If the errors are assumed to be distributed normally, the choice model takes the probit form. If the distribution of error terms is extreme value, the choice model with logit likelihood is obtained.
Take the binary probit model as an example, the choice model can be written as follows:

\[ y_i = 1 \quad \text{if} \quad x_i \beta + \epsilon_i > 0 \]
\[ \epsilon_i \sim N(0,1) \]  

(2.1)

To estimate the posterior distribution of $\beta$, it is necessary to integrate over a high dimensional parameter space.

\[ p(\beta \mid \{y_i\}, \{x_i\}) = \left( \int_{\Omega_t} \cdots \int_{\Omega_t} \epsilon_1 \epsilon_2 \cdots \epsilon_r d\epsilon_1 d\epsilon_2 \cdots d\epsilon_r \right) \cdot \pi(\beta) \]  

(2.2)

where $\Omega_t$ specifies the truncation region of $\epsilon_i$. For example, if $y_t=1$, $\Omega_t = \{ x_t \beta, \infty \}$. Otherwise, $\Omega_t = \{ -\infty, -x_t \beta \}$.

The estimation of the high dimensional integral can be avoided by introducing the latent variables $z_t$. The model can be rewritten as

\[ y_i = 1 \quad \text{if} \quad z_i > 0 \]
\[ z_i = x_i \beta + \epsilon_i \]
\[ \epsilon_i \sim N(0,1) \]  

(2.3)

Assume the prior distribution of $\beta$ is $N(\mu_0, \Sigma_0)$. The Gibbs sampler can be applied to simulate draws from the following conditional distributions of model parameters iteratively.

\[ [z_i \mid \text{else}] \propto [y_i \mid z_i][z_i \mid x_i \beta] \sim \text{Truncated Normal}(x_i \beta, 1) \]
\[ [\beta \mid \text{else}] \propto \prod_t [z_t \mid x_t \beta][\beta] \]
\[ \sim N \left( \left( \sum_{t=1}^T x_t'x_t + V_0^{-1} \right)^{-1} \left( \sum_{t=1}^T x_t'z_t + V_0^{-1} \mu_0 \right), \left( \sum_{t=1}^T x_t'x_t + V_0^{-1} \right)^{-1} \right) \]  

(2.4)

The data augmentation method has been applied to estimate latent variable models in marketing. For example, Edwards and Allenby (2003) propose a multivariate binomial
probit model to analyze multiple response data. The multivariate normal distribution is
treated as the latent construct so that standard multivariate analysis such as principle
components can be used to conduct exploratory analysis of survey data. Gilbride and
Allenby (2004) estimate a choice model that assumes consumers follow a discontinuous
decision process to make choice decision. The empirical result of this paper shows that
respondents use the conjunctive screening rules in a conjoint study. Notice that the latent
variable can be any unobserved construct of model, not necessarily latent utility. For
example, in a model with mixtures of normal components, the indicators of components
are viewed as augmented variables. Once the indicators are known, the observations can
be assigned to the normal component and other parameter estimations can be pursued
independently within each normal component (Rossi, Allenby, and McCulloch, 2005).

Error terms can also be augmented variables in a model since, in Bayesian paradigm,
error terms are unobservable and Bayesian treats all latent variables the same in a model.
For example, Allenby and Lenk (1994) analyze household purchase data with a logistic
normal regression model that allows cross-sectional and serial correlation in household
preference. The complexity of the error term structure requires generating draws of the
initial condition of error terms and autocorrelation parameters iteratively in the Gibbs
sampler. The initial condition of error terms is the augmented variable that facilitates the
estimation of other parameters in the model. Yang, Allenby, and Fennel (2002) treat error
terms as augmented variables in the estimation procedure of the model with the additive
heterogeneity distribution. It is necessary because the heterogeneity distribution assumes
that the same residual for a respondent is applied to all environmental fixed effect. After
obtaining the draw of a respondent’s residual, the coefficients for each respondent-
environment combination can be computed by adding the respondent’s residual and environmental fixed effect together. Zeithammer and Lenk (2005) use error augmentation to overcome the breakdown of the conjugacy between the covariance matrix and the inverted Wishart prior when there is a varying absent dimensions of the observations in a study. They suggest augmenting the absent residuals of a multivariate normal model, then estimating the full covariance matrix as if there are no absent dimensions.

Different from the application of error augmentation in marketing literature, the method of error augmentation developed in this dissertation is implemented with the procedure of checking the consistency of observed decisions and the decision rule that gives arise of the observed decision. The latent variables—the state variables in the first and the second essays and the latent utilities in the third essay—are generated first, then the error realization are retained to check if the decision rule defined in the model is consistent with the data when a candidate draw of parameter is generated.

To illustrate the error augmentation and the estimation algorithm proposed in this thesis, take the standard probit model shown in Equation (2.3) as an example. The model in Equation (2.3) can be expressed in terms of the errors

\[
\begin{align*}
    y_i &= 1 \quad \text{if} \quad \varepsilon_i > -x_i \beta \\
    \varepsilon_i &\sim N(0,1)
\end{align*}
\]

Assume that the prior of \( \beta \) is normally distributed with mean \( \mu_0 \) and variance \( \nu_0 \).

The model can be estimated by generating draws iteratively from the following conditional distributions
Note that $\Pi_{t}[y_t|\varepsilon_t,x_t\beta]$ is a product of indicator functions, the conditional distribution of the model parameter $\beta$ suggests a Metropolis-Hasting algorithm that retain the error realizations by $\varepsilon_t=z_t-x_t\beta$ for all $t$ and accept the candidate draw of $\beta$ (denoted by $\beta^{(n)}$) only if the latent utility $z_t$ given $\beta^{(n)}$ is consistent with the entire string of observation $y_t$. In other words, $\beta^{(n)}$ is accepted if $\varepsilon_t> x_t\beta^{(n)}$ and the purchase ($y_t=1$) is observed.

The alternative approach is to estimate the model by the new variant of error augmentation proposed in this thesis. The model in Equation (2.3) is

\begin{align*}
y_t = 1 & \text{ if } z_t > 0 \\
z_t = x_t\beta + \varepsilon_t \\
\varepsilon_t & \sim N(0,1)
\end{align*}

(2.7)

Given the same prior specification, the conditional distributions of the model parameters are

\begin{align*}
[z_t | \text{else}] & \propto [y_t | z_t][z_t | x_t\beta]\sim \text{Truncated Normal}(x_t,\beta,1) \\
\varepsilon_t = z_t - x_t\beta, \forall t \\
[\beta | \text{else}] & \propto \prod_t [y_t | x_t,\beta][\beta] \propto N(\beta | \mu_0, V_0) \cdot \prod_t [y_t | \beta, \varepsilon_t]
\end{align*}

(2.8)

Since $z_t$ is a function of $\beta$ and $\varepsilon_t$, and $\Pi_{t}[y_t|\varepsilon_t,x_t\beta]$ is a product of indicator functions, the conditional distribution of the model parameter $\beta$ suggests a Metropolis-Hasting algorithm that retains the error realizations by $\varepsilon_t=z_t-x_t\beta$ for all $t$ and accepts the candidate draw of $\beta$ (denoted by $\beta^{(n)}$) only if the latent utility $z_t$ given $\beta^{(n)}$ is consistent with the
entire string of observation $y_t$. In other words, $\beta^{(n)}$ is accepted if $z_t^{(n)} = x_t \beta^{(0)} + \varepsilon_t > 0$ and the purchase ($y_t = 1$) is observed.

As illustrated in this example, the binary probit model described in Equation (2.3) can be estimated by three different approaches—standard error augmentation (Equation (2.4)), the error augmentation (Equation (2.6)), and the new variant of error augmentation proposed in this thesis (Equation (2.8)). The estimation procedure of Equation (2.4) follows the standard data augmentation approach in which the latent utility $z_t$ is the augmented variable. Instead of generating the latent utility $z_t$, Equation (2.6) suggests treating the error $\varepsilon_t$ as the augmented variable and generating $\varepsilon_t$ from its conditional distribution. In this simple model, the error $\varepsilon_t$ can be generated directly from a distribution $[\varepsilon_t | \text{else}]$ because the full conditional specification is easy to specify. There is a direct correspondence between the observed data $y_t$ and the error $\varepsilon_t$. In other words, the observed choice $y_t$ is only related to one truncated normal distribution $TN(0, 1)$.

The estimation procedure of Equation (2.8) is specified according to the new approach of error augmentation proposed in this thesis. Different from Equation (2.4) and Equation (2.6), the error is not generated from a distribution directly, but computed from the latent utilities in Equation (2.8). This example shows that the proposed error augmentation can be a solution for estimating non-standard models that standard data augmentation and error augmentation cannot be applied. In this thesis, the development of error augmentation and the proposed estimation algorithm are provided in Chapter 3. The applications of the proposed approach are given in Chapter 3, 4, and 5.
3.1 Introduction

State-space models of behavior assume that demand is related to parameters and latent variables that evolve through time. Examples include models of purchase timing and quantity that are affected by unobserved household inventories, and models of consumer learning where consumer preference for product attributes is determined, in part, by advertising and consumption experience. When these models of behavior are developed within an economic framework of utility maximization, analysis is complicated by the fact that utility function parameters and common error terms can be present in the observation equation, where behavior is related to marginal utility, and the state equation, where the evolution of the arguments of the utility function is described.

The presence of shared parameters and common error terms invalidate the use of standard methods of estimation such as the Kalman filter (Meinhold and Singpurwalla 1983). While algorithms are available for dealing with nonlinear and non-Gaussian state-space models (e.g., Kitagawa 1987, West and Harrison 1997, de Jong and Shepard 1995), they rely on transformations that yield an approximate Guassian likelihood within a
normal-normal model that do not accommodate these characteristics. Consider, for example, a choice model involving latent household inventory of a product. If utility is a function of inventory such that diminishing marginal returns are present, then observed demand and the marginal utility of purchasing additional units of the product will depend on latent inventory and parameters of the utility function. Moreover, the same error realizations associated with the utility function will be present in the expression for marginal utility.

The purpose of this paper is to develop a hierarchical Bayes approach to estimating state-space models of economic behavior by employing a new variant of data augmentation. With data augmentation, realizations of the state variable are simulated as part of the estimation procedure and are used to simplify the evaluation of the likelihood function. Typical applications of data augmentation involve generating latent variables, such as utilities, associated with models of choice. The presence of common parameters in the observation and state equations, and shared error realizations, requires the development of a new algorithm to avoid the evaluation of the likelihood. The algorithm has application to a wide class of micro-economic data and models.

The organization of the paper is as follows. In the next section problems encountered in estimating state-space models of economic behavior are discussed, and the proposed Bayesian approach is introduced. Section 3.2 illustrates use of the model in the context of a direct marketing problem. Data description and parameter estimates are provided in section 3.3, and section 3.4 provides a discussion of the results. Concluding comments are offered in section 3.5.
3.2 State-space Models for Economic Behavior

State-space models comprise an observation equation that relates an observed behavior to a latent variable, whose temporal variation is described by a state equation. In marketing, observed behavior is often discrete, and a general parametric representation of the state-space model for economic behavior is specified as:

Observation Equation: \( y_t = k \) if \( \delta(s_t, \beta) \) is true \hspace{1cm} (3.1)

State Equation: \( s_t = F(s_{t-1}, \beta, x_t, y_{t-1}) + \varepsilon_t \) \hspace{1cm} (3.2)

where a discrete observation, \( k \), corresponds to a decision rule \( \delta(s_t, \beta) \) with state variable \( s_t \) that evolves through time stochastically. The parametric structure of the decision rule \( \delta(.) \) and evolution process \( F(.) \) arise naturally from underlying theory associated with the study, and may describe either linear or non-linear relationships among variables. Note that three important aspects of this formulation: i) the parameter vector \( \beta \) is present in both equations, ii) the dependence of \( s_t \) on \( s_{t-1} \) in the state equation results in an auto-correlated process in the presence of the error term, \( \varepsilon_t \), and iii) the response variable in the observation equation is a discrete realization of a continuous latent variable and shared parameters. Approximating filtering and smoothing algorithms suggested in the literature (e.g., Meinhold and Singpurwalla 1983; West and Harrison 1997; Carlin, Polson, and Stoffer 1992; de Jong and Shephard 1995; Carter and Kohn 1994) provide solutions to estimate non-linear and non-Gaussian state space models. These algorithms, however, cannot be used to estimate state-space models with a discrete observation equation containing shared parameters because these three properties of state-space model described above lead to a non-standard distribution of likelihood.
Simplified versions of our state-space model have been used in the marketing literature, typically by assuming the state equation (e.g., inventory) evolves deterministically, or follows a simple process. Deterministic updating can be found in models where the stochastic element, $\epsilon_t$, is assumed part of the observation equation and whose effect does not propagate through time (Gonul and Srinivasan 1996; Sun, Nelsin and Srinivasan 2003). If, however, the deterministic part of the state-equation is misspecified, the implied error distribution at the observation equation is usually intractable. Including an error term in the state equation leads to a more robust model specification.

Discrete choice models typically assume that the state variable (i.e., utility) is linear in the parameters with no carry-over, and when carry-over is present, it is specified so that there are tractable updating equations for $F(.)$ (e.g., Allenby and Lenk 1994; Erdem and Keane 1996; Seetharaman, Ainslie, and Chintagunta 1999). The assumption of linear utility implies that marginal utility is constant and does not dependent on model parameters. The proposed model therefore represents a generalization application to situations where the utility function is non-linear and the evolution of the state variable is less restrictive.

To motivate the need for the proposed estimator, consider a consumer who is recruited into a membership program (e.g., classical music club, subscription to repair books, etc.) where they periodically receive offers for evaluation. The consumer is assumed to hold an unobserved inventory ($s_t$) of the good being sold, which is potentially depleted over time ($t$). The consumer elects to make a purchase when the marginal utility
of the offer is sufficiently high. When a purchase is made, the inventory level increases by an amount, $\beta$, to be estimated. A state-space representation of this process is:

**Observation Equation:**

$$ y_i = \begin{cases} 
1 & \text{if } (s_t + \beta)^\rho - s_t^\rho \geq \gamma \\
0 & \text{otherwise} 
\end{cases} \quad (3.3) $$

**State Equation:**

$$ s_t = \phi s_{t-1} + \beta y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim N(0,1) \quad (3.4) $$

where $\beta$ is a common parameter that represents the inventory equivalent of the good, $\rho$ is a parameter that reflects diminishing marginal returns to holding inventory ($0 \leq \rho \leq 1$), $\phi$ is a parameter that reflects the depletion of inventory ($0 < \phi < 1$), and $\gamma$ is a threshold, which, if exceeded, results in the purchase of an offering. Customers decide to keep an offering if its incremental value is sufficiently high, and will return an offering if the incremental value is lower than the threshold. The offerings are viewed as adding to the consumer’s inventory of the product category, which are subject to diminishing marginal returns.

Two challenges exist in estimating the state-space model described by Equations (3.3) and (3.4). The first challenge is in dealing with the autocorrelation of the state variable, $s_t$, which is of nonstandard form because of the covariate $y_{t-1}$. As discussed by Keane and Wolpin (1994), state space models, in general, suffer from the need to compute high-dimensional integrals in evaluating the likelihood of the observed data. Despite the simple binary nature of the outcome variable, the choice probability involves computing a $t$ dimensional integral because the stochastic shocks are a function of $\{\varepsilon_t, \ldots, \varepsilon_t\}$.

The computational complexity associated with state-space models is addressed with the Bayesian method of data augmentation (Tanner and Wong, 1987). Data augmentation avoids the need to evaluate high-dimensional integrals by treating the state parameter as
an unobserved latent variable in the model, with estimation proceeded by conditioning on realizations of the state variable, $s_t$, in a Markov chain. The ability to condition on realizations of the state parameter leads to a deterministic observation equation (Equations (3.1) or (3.3)) resembling an indicator function, where state parameters are either consistent or not consistent with the observed data. The state-space is then navigated using properties of the Markov chain, instead of attempting to marginalize the likelihood function by integrating out the latent variable. The presence of serial correlation requires a non-standard method of data augmentation, which is described below, and which can be applied to complicated observation equations.

The second challenge is the ability to empirically identify all the model parameters. The data are not simultaneously informative, for example, about the location of the state variable, the discount parameter, $\rho$, and the threshold parameter, $\gamma$. Equivalent realizations of the outcome variable $\{y_t\}$ can arise with alternative threshold and discount parameters by changing the location of the state variable. A method of evaluating the likelihood function is proposed so that identification issues can be investigated.

3.2.1 Model Estimation

The discussion of the estimation algorithm is motivated by considering the Bayesian method of data augmentation applied to the binomial probit model (Albert and Chib 1993). While standard methods of data augmentation can be used to estimate parameters of the probit model, the new method of data augmentation for this model is introduced to illustrate differences. However, note that standard models cannot be used to estimate the
model described by Equations (3.3) and (3.4) because of the presence of common parameters and common auto-correlated errors.

The method of data augmentation involves the introduction of latent variables into a hierarchical model to simplify computation. The augmented variable for the binomial probit model is a latent continuous variable, which, if positive, indicates that the binomial realization is equal to one:

\[
y_i = \begin{cases} 
1 & \text{if } z_i \geq 0 \\
0 & \text{if } z_i < 0
\end{cases} \tag{3.5}
\]

\[
z_i = \alpha + \varepsilon_i, \quad \varepsilon_i \sim N(0,1), \quad \text{for } t = 1,2,\ldots,T \tag{3.6}
\]

The associated hierarchical representation of the model is:

\[
[y_i | z_i] \tag{3.7}
\]

\[
[z_i | \alpha] \tag{3.8}
\]

\[
[\alpha] \tag{3.9}
\]

where \([y_i|z_i]\) is an indicator function equal to one if \(z_i \geq 0\), and \([z_i|\alpha]\) is distributed normal with mean \(\alpha\) and variance one, and \([\alpha]\) is a prior distribution for \(\alpha\). Estimation can be carried out using Gibbs sampling by generating draws from the full conditional distribution of model parameters \(\{z_i\}\) and \(\alpha\):

\[
[z_i | else] \propto [y_i | z_i][z_i | \alpha] \sim \text{Truncated Normal}(\alpha,1) \tag{3.10}
\]

\[
[\alpha | else] \propto \prod_t [z_i | \alpha][\alpha] \sim \text{Normal}(\overline{z_i},1/T) \tag{3.11}
\]

for the prior on \(\alpha, [\alpha]\), assumed uniform over a large region.
An alternative approach to Bayesian estimation of the binomial probit model, which is required for the state-space model, is to treat the error term, $\varepsilon_t$, as the augmented variable instead of $z_t$. The model is defined as:

$$
\begin{align*}
    y_t = \begin{cases} 
    1 & \text{if } \alpha + \varepsilon_t \geq 0 \\
    0 & \text{if } \alpha + \varepsilon_t < 0
    \end{cases} \\
    \varepsilon_t \sim N(0,1), \text{ for } t = 1, 2, \cdots, T
\end{align*}
$$

(3.12)

(3.13)

or, hierarchically, as:

$$
[y_t | \varepsilon_t, \alpha] \quad (3.14)
$$

$$
[y_t | \varepsilon_t] \quad (3.15)
$$

$$
[\alpha] \quad (3.16)
$$

where $[y_t | \varepsilon_t, \alpha]$ is an indicator function and $[\varepsilon_t]$ is distributed normal with mean zero and unit variance. The conditional distribution of model parameters is:

$$
[y_t | \varepsilon_t, \alpha] \propto [y_t | \varepsilon_t, \alpha] [\varepsilon_t] \sim N(0,1) \cdot I(y_t | \varepsilon_t, \alpha) \quad (3.17)
$$

$$
[\alpha | \text{else}] \propto \prod_t [y_t | \varepsilon_t, \alpha] [\alpha] \quad (3.18)
$$

where the conditional distribution for $\alpha$ involves the product of indicator functions times the prior distribution. Sampling from the conditional distribution of $\alpha$ is straightforward with the Metropolis-Hastings algorithm. For example, a random-walk chain would involve generating a new draw from a previous draw plus normal error, $\alpha^{(n)} = \alpha^{(p)} + \Delta \alpha$, and accepting the new draw with probability $\kappa$:

$$
\kappa = \min \left( \frac{\prod_t [y_t | \varepsilon_t, \alpha^{(n)}] [\alpha^{(n)}]}{\prod_t [y_t | \varepsilon_t, \alpha^{(p)}] [\alpha^{(p)}]} , 1 \right) \quad (3.19)
$$
where $\Pi_t[y_t|\varepsilon_t,\alpha]$ is a product of indicator functions. A candidate value, $\alpha^{(n)}$, is never accepted unless it the quantity $\alpha^{(n)} + \varepsilon_t$ is consistent with $y_t$ for $t = 1, \ldots, T$. If $\alpha^{(n)}$ is consistent with the observed data, then it is accepted with probability determined by the prior distribution $[\alpha]$. It is important to note that use of Equation (3.19) requires the initial value of $\alpha$ to be associated with a product of indicator functions equal to one, i.e., a value in the valid region, so that the denominator is non-zero.

However, the estimation procedure illustrated in Equations (3.17)-(3.18) cannot be applied to state-space models with auto-correlated state variables (e.g., Equation (3.4)). Since there is no correspondence between the observed choice $y_t$ and the error $\varepsilon_t$ when the autocorrelation among state variables presents, the conditional distribution of the error $\varepsilon_t$ is difficult to specify and cannot be generated from a distribution directly. Therefore, the new estimation procedure is proposed in this paper to deal with the state-space models with common parameters present in the observation and state equation, and common error realizations.

Take the binomial probit model illustrated in Equation (3.5)-(3.6) as an example. The new estimation approach suggest

(1) generating the latent variable $z_t$ from the conditional distribution

$$[z_t | \text{else}] \propto [y_t | z_t][z_t | \alpha] \sim \text{Truncated Normal}(\alpha, 1) \quad (3.20)$$

(2) retaining the error realization by

$$\varepsilon_t = z_t - \alpha \quad (3.21)$$

(3) generating the draws of $\alpha$ from its conditional distribution

$$[\alpha | \text{else}] \propto \prod_t [y_t | \varepsilon_t, \alpha][\alpha] \quad (3.22)$$
Note that the only difference between the estimation procedure of Equation (3.17)-(3.18) and the estimation procedure of Equation (3.20)-(3.22) is whether the error $\varepsilon_t$ or $z_t$ are generated from a distribution.

The advantage of the proposed approach is that it is better able to deal with complicated model structures such as the state-space model described by Equations (3.3) and (3.4) that cannot be estimated with standard data augmentation approach. The disadvantage, however, is that convergence occurs at a much slower rate because the Markov chain is not optimally exploiting the distributional structure of the hierarchy.

Figure 3.1 illustrates this tradeoff for $\alpha = 0.50$ and $T = 100$, an extreme example in that most marketing data are characterized by shorter purchase histories at the individual level. The figure displays time series plots for 50,000 iterations of the standard algorithm (Equation (3.10) – (3.11)) and the proposed algorithm (Equations (3.20) – (3.22)). The standard algorithm converges immediately to the true posterior distribution, whereas convergence of the proposed algorithm is much slower and exhibits higher autocorrelation. The extent of autocorrelation is directly related to the length of the data ($T$), with shorter histories associated with smaller autocorrelation.

The proposed approach can be employed to estimate the state-space model described by Equations (3.3) and (3.4), which cannot be estimated with standard methods. The hierarchical representation for the model is:

$$[y_t | s_t, \beta, \rho, \gamma]$$ \hspace{1cm} (3.23)

$$[s_t | s_{t-1}, y_{t-1}, \beta, \phi]$$ \hspace{1cm} (3.24)

$$[\beta]$$ \hspace{1cm} (3.25)
The key insight to our method is in recognizing there is correspondence between the state variable, \( s_t \), and the error term, \( \varepsilon_t \). The estimation procedure is begun by generating draws of the state variables, conditional on all other model parameters, and from these, back out realizations of the error by using the state equation. Once the error realizations are obtained, candidate values of other parameters, such as \( \phi \) or \( \beta \) in Equation (3.4), are used to construct new values of the state variables.

Note that, instead of drawing \( \varepsilon_t \) from its conditional distribution, the proposed approach suggest generating draws of the state variables, \( s_t \), and solving for \( \varepsilon_t \) because the conditional distribution of the error term is intractable. Assuming that the unconditional distribution of \( \varepsilon_t \) is normal with unit variance, the distribution of the state variable \( s = (s_1, s_2, s_3, \ldots, s_T)' \) in Equation (3.4) can be shown to be normally distributed:

\[
s \sim \text{Normal}(\mu, \Sigma) ;
\]

\[
\mu = \begin{pmatrix}
\mu_1 = \phi s_0 + \beta y_0 \\
\mu_2 = \phi \mu_1 + \beta y_1 \\
\mu_3 = \phi \mu_2 + \beta y_2 \\
\vdots \\
\mu_T = \phi \mu_{T-1} + \beta y_{T-1}
\end{pmatrix}
\]

\[
\Sigma = \begin{bmatrix}
1 & \phi & \phi^2 & \cdots & \phi^{T-1} \\
\phi & \phi^2 + 1 & \phi (\phi^2 + 1) & \cdots & \phi^{T-2} (\phi^2 + 1) \\
\phi^2 & \phi (\phi^2 + 1) & \phi^4 + \phi^2 + 1 & \cdots & \phi^{T-3} (\phi^4 + \phi^2 + 1) \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\phi^{T-1} & \phi^{T-2} (\phi^2 + 1) & \phi^{T-3} (\phi^4 + \phi^2 + 1) & \cdots & \phi^{T-2} (\phi^4 + \phi^2 + 1) + \phi^2 + 1
\end{bmatrix}
\]

where \( s_0 \) is the initial value of the state variable. Realizations of the state variable are obtained by generating draws from the full conditional distribution of \( s_t | s_{-t} \) where "-t"
denotes "all elements except \( t \)." The conditional distribution is distributed truncated normal:

\[
[s_t | s_{-t, \text{rest}}] \sim \text{Normal}[\mu_{st} + \Sigma_{t,t}^{-1} \Sigma_{t,\text{rest}} (s_{-t} - \mu_{-t}), \tau_{t,t}] \cdot I(y_t)
\]  

(3.30)

where \( \mu_{st} \) is the mean corresponding to the state variable \( s_t \), \( \Sigma_{t,t} \) be the \( t \)th column of covariance matrix \( \Sigma \) excluding the \( t \)th element from this vector, \( \Sigma_{-t,t} \) be a matrix after removing the \( t \)th column and the \( t \)th vector from the covariance matrix \( \Sigma \), \( \tau_{t,t} \) is the \( t \)th element of \( 1/\text{diag}(\Sigma^{-1}) \) and is equal to the conditional variance (see McCulloch and Rossi 1994), and \( I(y_t) \) is an indicator function equal to one if Equation (3.3) is true for the draw of \( s_t \). Once draws of \( \{s_t, \ t = 1, \ldots, T\} \) are obtained, corresponding realizations of the error terms \( \{\epsilon_t, \ t = 1, \ldots, T\} \) are easily computed.

Estimation proceeds by generating draws of the other model parameters using the computed values of the error term and the Metropolis-Hastings algorithm. Consider, for example, the autoregressive parameter \( \phi \). A candidate draw of \( \phi \) using a random walk chain is obtained from the previous draw, \( \phi^{(n)} = \phi^{(p)} + \Delta \phi \), and used to construct a new realization of the state vector using the state equation \( s_t^{(n)} = \phi^{(n)} s_{t-1}^{(n)} + \beta y_{t-1} + \epsilon_t \), accepted with probability \( \kappa \):

\[
\kappa = \min \left( \frac{\prod_t [y_t | s_t^{(n)}, \beta, \gamma][\phi^{(n)}]}{\prod_t [y_t | s_t^{(p)}, \beta, \gamma][\phi^{(p)}]}, 1 \right)
\]  

(3.31)

Estimation of other model parameters proceeds in a similar way.
3.2.2 Model Identification

Not all parameters in Equations (3.3) and (3.4) are statistically identified. The observation equation relates an observed binary outcome to a state variable \((s_i)\), an effect size \((\beta)\), a discount parameter \((\rho)\), and a threshold \((\gamma)\). In contrast, in the standard probit model (Equation (3.5)), the binary outcome is related to just one latent variable \((z_i)\). The presence of the state equation allows for identification of an additional model parameters, and the illustration of our model is proceeded by assuming that \(\beta\) and \(\phi\), the autoregressive coefficient in the state equation, are the parameters of interest.

To demonstrate the identification problem among model parameters, the contour plots displayed in Figure 3.2 are provided to visualize the relationship among \(\rho\), \(\gamma\), \(s\), and \(\beta\) in the decision rule \((s + \beta)^\rho - s^\rho = \gamma\). The autocorrelation coefficient \(\phi\) is not included in this analysis because \(s\) is the function of \(\phi\) in Equation (3.4). The contours displayed in Figure 3.2(a) are the values of \(\beta\) and \(s\) that solve the observation equation \((s + \beta)^\rho - s^\rho = \gamma\), for the threshold parameter, \(\gamma\), taking on values of 0.35, 0.40 and 0.45, and the discount parameter, \(\rho\), equal to 0.50. The contours displayed in Figure 3.2(b) are the values of \(\beta\) and \(s\) that solve the observation equation \((s + \beta)^\rho - s^\rho = \gamma\), for the discount parameter, \(\rho\), taking on values of 0.6, 0.7 and 0.8, and the threshold parameter, \(\gamma\), equal to 0.4. Both plots in Figure 3.2 reveal that there are multiple solution of \((\beta, s)\) given different values of \(\rho\) and \(\gamma\). Since not all model parameters in the model are identified, their interpretation becomes more complex, and the evaluation of marketing policies and events requires the estimation of effect sizes in terms of choice probabilities.
Model identification in state-space models is not always straightforward, and it is useful to be able to evaluate the likelihood of data to determine which parameters are identifiable. The likelihood corresponds to a region of the underlying error distribution revealed by the choice data. The likelihood can be evaluated by simulating draws from the error term in Equation (3.2) or (3.4), and counting the proportion of times the observation equation is true. For our example,

1. Generate $\epsilon_i \sim N(0,1)$ for $i=1,\ldots,N$ and $t=1,\ldots,T$

Construct a set of realizations of the state variable for $s_i$ conditional on the lagged state variable $(s_{i,t-1})$ and other model parameters $(\beta, \phi)$ obtained in the $k$th iteration of the Markov chain:

$$s_i^t = \phi^t s_{i-1} + \beta^t y_{i,t-1} + \epsilon_i^t, \text{ for } i = 1,2,\ldots,N; t = 1,2,\ldots,T$$ \hspace{1cm} (3.32)

2. Determine the frequency the simulated state variable, $s_i$, satisfies the observation equation. The frequency serves as an estimate of the likelihood:

$$\text{Pr}(y_i \mid \phi, \beta) = \frac{1}{N} \sum [y_i \mid s_i^t]$$ \hspace{1cm} (3.33)

3. The joint likelihood of the data, evaluated at the current realization of the Markov chain, is equal to the product of the choice probabilities for each observation:

$$\text{Pr}(y_1, y_2, \ldots, y_T \mid \phi, \beta) = \text{Pr}(y_T \mid y_1, y_2, \ldots, y_{T-1}, \phi, \beta) \cdot \text{Pr}(y_{T-1} \mid y_1, y_2, \ldots, y_{T-2}, \phi, \beta) \cdot \ldots \text{Pr}(y_2 \mid y_1, \phi, \beta) \cdot \text{Pr}(y_1 \mid \phi, \beta)$$ \hspace{1cm} (3.34)
3.2.3 Simulation Study

To illustrate convergence of the proposed algorithm, 100 realizations of the model described in Equations (3.3) and (3.4) are simulated assuming that $s_0 = 5.0$, $\rho = 0.5$, $\phi = 0.7$, $\gamma = 0.4$ and $\beta = 2.0$. The error term is assumed normal with unit variance. Figure 3.3 provides time series plots of the estimated parameters $\rho$ and $\beta$, which are seen to converge to the true model parameters indicated by the horizontal lines. Note that, if the data were estimated with a model that incorrectly specified the covariance matrix ($\Sigma$) as diagonal, the algorithm would fail to converge. The lower truncation of the posterior distribution of $\beta$ is due to the effect of the initial value $s_0$ on the initial purchase decision $y_0$.

An important aspect of the estimation procedure is in obtaining valid starting values of the parameters so that the previous draws in Equation (3.19) (e.g., $\phi^{(p)}$) are valid, or are associated with a non-zero values of the likelihood. This can become difficult when the number of observations, $T$, is large because the product of indicator functions associated with the observation equation (see Equation (3.22)) can limit the support of posterior distribution of model parameters. In this case, it is often useful to allow the initial value of the state variable, $s_0$, to initially take on a value that is different from its true value, using a grid search procedure to identify valid starting values of other model parameters, and gradually move $s_0$ toward its true value. Such a procedure was used in the simulation, where initial values of the parameters where $\phi = 0.4$, $\beta = 1.0$, and $s_0 = 1.0$. At iteration 2 million, the value of $s_0$ reached the true value of 5.0, and convergence occurred quickly thereafter. In the empirical application discussed below, $s_0$ is treated as a parameter and estimate its posterior distribution, speeding convergence of the chain.

26
3.3 Direct Marketing Application

Data were obtained from a direct marketing organization specializing in continuity programs selling items related to a particular theme. Examples include book series about medicine, home repair, children’s books and recordings from a particular time period (e.g., Songs of the 70's). Customers were recruited to the program through television commercials to purchase the initial item in the series, with the understanding that the customer would receive additional items for their consideration in a series of future mailings. Customers in the program continued to receive additional items provided they paid for the item shipped, or return the item they did not want. The inter-arrival time between mailings is four weeks. There are typically more than 30 items in a particular series, although few of the customers purchase all items or even remain in the program to its conclusion. In fact, the average number of shipments to customers enrolled in the series averages between five and seven, depending on the specifics of the program, and frequently only two or three actual purchases are made after the initial item.

A random sample of 250 customers was obtained from company records. Figure 3.4 displays descriptive aspects of the data. The top portion of Figure 3.4 displays the number of shipments and purchases of the items shipped. There were 32 distinct items offered for sale, with the first item (i.e., item 1) received by all customers as part of the introductory offer. Items were selected at random from company inventory for shipment to customers, and were not offered in a particular pattern (e.g., item 1, then item 2 and so on). The bottom portion of the figure displays the distribution of purchase histories, denoted $T_j$, where $j$ is a customer index. The minimum customer longevity is two periods, indicating the decision to resign after accepting the introductory offer in period one and either
accepting or declining an item offer in the second period. The sample statistics match closely the corresponding statistics from the universe of customer records maintained by the firm.

The model in Equations (3.3) and (3.4) is expanded to accommodate the customer resignation decision and customer heterogeneity. There are two observation equations for our state-space model, one for the purchase decision \( y \) and one for the resignation decision \( x \). Resignation from the continuity program (i.e., \( x = 1 \)) is assumed to occur when inventory reaches a level where the marginal utility of an additional offering is low.

**Observation Equations:**

\[
\begin{align*}
y_{t,j} &= \begin{cases} 1 & \text{if } (s_{t,j} + \omega_{k(t),j})^p - s_{t,j}^p \geq \gamma \\ 0 & \text{otherwise} \end{cases} \quad (3.35) \\
x_{t,j} &= \begin{cases} 1 & \text{if } \rho(s_{t,j} + \omega_{k(t),j} \cdot y_{t,j})^{p-1} \leq \delta \\ 0 & \text{otherwise} \end{cases} \quad (3.36)
\end{align*}
\]

**State Equation:**

\[
s_{t,j} = \phi_j s_{t-1,j} + \omega_{k(t-1),j} y_{t-1,j} + \varepsilon_{t,j}, \quad \varepsilon_{t,j} \sim N(0,1) \quad (3.37)
\]

where \( \varepsilon_{t,j} \) is distributed standard normal, and \( \omega_{k(t),j} \) the effect of \( k \)th item received by customer \( j \) in period \( t \), is model as a item effect, coupled with customer heterogeneity:

\[
\omega_{k(t),j} = \alpha_{k(t)} \cdot \beta_j, \quad \text{for } t = 0,1,\cdots,T_j \text{ and } j = 1,2,\cdots,J \quad (3.38)
\]

\[
\alpha_{k(t)} = \exp(\alpha_{k(t)}^*) \quad (3.39)
\]

\[
\beta_j = \exp(\beta_j^*); \quad \beta_j^* \sim N(0,\sigma_\beta^2) \quad (3.40)
\]

The customer and item effects are parameterized with exponential functions to ensure the effect is positive. The item effects \( \alpha \) are treated as fixed effects, while the customer effects are model as random-effects with mean zero. Finally, the initial condition and autoregressive parameters are parameterized to appropriate regions.
Equation (3.35) is identical to observation equation (Equation (3.3)) where consumers make purchases when the marginal benefit of accepting the offer is sufficiently high. Equation (3.36) models the resignation decision after making the purchase decision, $y_{t,j}$. The state equation (Equation (3.37)) is identical to Equation (3.4) except for the presence of a more complicated specification for the item effect-size, $(\omega_{k(t),j})$, that includes random effects and range restrictions.

The presence of the additional observation equation (Equation (3.36)) for the resignation decision allows identification of additional model parameters. In addition to estimating the item effect sizes $(\omega_{k(t),j})$ and autoregressive parameters $(\phi_j)$ as in the simulation study above, the resignation data allows estimation of the initial inventory level ($s_0$).

The bivariate data $\{y_{t,j}, x_{t,j}\}$ can take on four different outcomes: purchase and no resignation ($y_{t,j}=1, x_{t,j}=0$), purchase and resign ($y_{t,j}=1, x_{t,j}=1$), no purchase and no resignation ($y_{t,j}=0, x_{t,j}=0$), and no purchase and resign ($y_{t,j}=0, x_{t,j}=1$). Since the data are assumed to be related to the same set of state variables, $\{s_{t,j}\}$, the observation equations impose two points of truncation when generating the augmented variable. Let $T_{y_{t,j}}$ denote the truncation point associated with the purchase decision, obtained by solving for the value of $s_{t,j}$ in Equation (3.35) such that the left and right sides of the equation are equal. Similarly, let $T_{x_{t,j}}$ denote the truncation point associated with the resignation decision (Equation (3.36)). Then, the data are related to the latent variable $s_{t,j}$ as follows:

$$s_{0,j} = \exp(s_{0,j}^*) ; \quad s_{0,j}^* \sim N(\bar{s}_0, \sigma_{s_0}^2)$$

$$\phi_j = \frac{\exp(\phi_j^*)}{1 + \exp(\phi_j^*)} ; \quad \phi_j^* \sim N(\bar{\phi}, \sigma_{\phi}^2)$$

(3.41)  

(3.42)
Algorithms for estimating the proposed model are provided in Appendix A. Proper, but diffuse priors are assumed for all model parameters. The estimates reported in the next section are based on a Markov chain run for 30,000 iterations, the last 10,000 iterations used for generating parameter estimates. Convergence was checked by starting the chain from multiple initial values, and by inspection of time series plots.

### 3.4 Empirical Results

Parameter estimates are reported in Table 3.1. As discussed above, the estimates are conditional on the values selected for the non-identified parameters. The parameter $\rho$ that reflects diminishing returns to scale was set to 0.5, the observation cutoff parameter of purchase decision, $\gamma$, was set to 0.15, and the observation cutoff parameter of resigning decision, $\delta$, was set to 0.15. These values are selected because they do not overly restrict the truncation points associated with the observation equation (e.g., too large in magnitude). Note that policy implications of the model reported below do not depend on these values. Moreover, the conditional nature of the estimates implies that the absolute magnitude of many of the coefficients is not directly interpretable, although the relative magnitudes are comparable. For example, while the absolute magnitude of the item effect $\omega_{h(0,j)}$ is dependent on the rate of diminishing returns implied by $\rho$, the relative magnitude...
of $\omega_{k(t),j}$ among customers ($j$) and offered items ($k$) is comparable. The associated customer ($\beta_j = \exp(\beta^*_j)$) and item effects ($\alpha_{k(t)} = \exp(\alpha^*_{k(t)})$) are displayed in Figure 3.5, and are seen to have posterior mass away from zero.

Figure 3.6 displays the relationship between the posterior mean of the item effects ($\alpha_{k(t)}$) and the observed acceptance rate for each book. The Pearson correlation coefficient is 0.729, indicating close agreement between the data and model estimates. In general, books with higher acceptance rates have larger estimated effect sizes, as expected.

Figure 3.7 displays the distribution of autocorrelation coefficients $\phi_j$. The distribution is centered very near 1.0, implying that the state variable does not depreciate much over time. This corresponds to the durable nature of the product items in the dataset, and the relatively short inter-offer time (i.e., 4 weeks) of the mailings. It is interesting to note that, in addition to the mean of the distribution being close to 1.0, the variability of the coefficient is small, with only a small number of customer coefficients with coefficients that indicate significant depletion of the state variable between mailings.

Figure 3.8 displays the relationship between the posterior mean of the autocorrelation coefficient ($\phi_j$) and the longevity of each customer ($T_j$). A negative association exists between the autocorrelation coefficient and customer longevity, indicating that customers who tend to stay in the club longer are associated with smaller value of $\phi$, or consume their inventory at a faster rate.

Finally, Figure 3.9 displays posterior estimates of initial states, $s_0$, for each customer. The distribution has a concentration of mass at zero, which is associated with customers with longer longevity ($T_j$) and smaller autoregressive coefficients ($\phi_j$) than other
customers, indicating that it takes longer for these individuals to build up their inventory to the point where they resign from the continuity program.

3.5 Discussion

The results indicate that the relatively short longevity of many customers is due to the high autocorrelation of the state variable, where customer inventory is not depleted sufficiently fast in light of the frequency of new offers arriving by mail. This raises the possibility that longer periods of item inter-arrival may lead to longer customer longevity and greater sales. The effect of increasing the inter-arrival time on sales is investigated by computing the expected sales for each customer conditional on posterior estimates of their model parameters, and exploring the relationship to changes in the autocorrelation coefficient. An autocorrelation of $\phi$ implies that $(1-\phi)$ of the offering is consumed each period (i.e., every 4 weeks). Doubling the inter-arrival time implies consumption of $2(1-\phi)$ units, or an autocorrelation coefficient of $(1-2(1-\phi))$, and so on. In the plots below, it is assumed that the time value for money is ten percent and demand is computed over a five year horizon.

Figure 3.10 plots expected demand against changes in the inter-arrival time of the offerings. For each inter-arrival time investigated, the associated change in the autocorrelation coefficient is determined for each customer, and expected demand is computed by simulating realizations of the state equation and associated realizations of the observation equations. The plot indicates increasing the inter-arrival time from one month to two months can have maximum discounted volume of expected demand.
3.6 Concluding remarks

In this paper a new approach is proposed to estimating state-space models of economic behavior. The approach is based on the Bayesian method of data augmentation where latent realizations of the state variables are simulated according to Bayes rule, and are used construct corresponding realizations of the error terms. These error terms are then used to reconstruct state variables for different values of the parameters, thus simplifying the high-dimensional analysis associated with the estimation of latent variables models. Direct marketing data are used to illustrate the model, where the model is used to investigate potential gains from increasing the inter-arrival time of offers in a membership club.

Marketing data is limited in its ability to reveal latent processes associated with consumer behavior. Bayesian methods offer an ideal approach to inference because they have known finite-sample properties. While gains have been made in estimating models based on random utility theory, there is a growing awareness that marketing data reflect a complicated process, often involving interventions and stochastic components that vary over time. Researchers have documented, for example, the effect of variety seeking and state dependence on consumer choice (McAlister 1982; Kahn, Kalwani and Morrison 1986; Erdem 1996), the presence of structural heterogeneity in choices (Kamakura, Kim and Lee 1996; Allenby, Leone and Jen 1999), and the presence of motivational components of behavior that arises from the context of consumption (Yang, Allenby and Fennell 2002). This literature supports the view that the unit-of-analysis is not the individual consumer – it is the individual engaged in an occurrence of an activity, or at a specific point in time. The difficulty of employing models of behavior for an activity-
occasion is that limited data exists at the disaggregate level. State-space models of behavior, in which latent variables, instead of observed data, are assumed to influence consumer behavior therefore offers a fruitful domain for research, and the proposed approach offers a flexible framework for analyzing such behavior.

Many extensions to the basic model, and applications of state-space models to marketing data, are possible. Examples include models of consumer search and learning that where the state variable corresponds to information and beliefs of the consumer, and models where the state equation is described by multiple states (e.g., conscious and pre-conscious cognitive processing). The proposed Bayesian estimator offers a general approach to estimating many such extension
Figure 3.1 Markov chain realizations of $\alpha$: (a) New algorithm; (b) Standard algorithm
Figure 3.2 Identification Analysis for latent inventory (s) and effect size (\(\beta\)): (a) Contours of equal probability of \(s\) and \(\beta\), for \(\rho = 0.5\); (b) Contours of equal probability of \(s\) and \(\beta\), for \(\gamma = 0.4\)
Figure 3.3 Markov chain realizations of model parameters for simulation study: (a) autoregressive parameter ($\phi$); (b) Effect-size parameter ($\beta$)
Figure 3.4 Descriptive statistics: (a) Item shipments and purchases; (b) Distribution of household purchase history ($T_j$)
Figure 3.5 Posterior distribution of customer and item effects: (a) customer effects ($\beta_j$); (b) Item effects ($\alpha_k$)
Figure 3.6 Acceptance rates versus items effects ($\alpha_k$)
Figure 3.7 Posterior distribution of autocorrelation coefficients (\(\phi\))
Figure 3.8 Customer longevity ($T_j$) versus autocorrelation coefficient ($\phi$)
Figure 3.9 Posterior distribution of initial state ($s_0$)
Figure 3.10 Expected demand for offering inter-arrival times
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Posterior mean</th>
<th>Posterior standard deviation</th>
<th>95% HPD Credible Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_2$</td>
<td>-0.0223</td>
<td>0.0103</td>
<td>(-0.0431, -0.0031)</td>
</tr>
<tr>
<td>$\alpha_3$</td>
<td>-0.0394</td>
<td>0.0106</td>
<td>(-0.0593, -0.0176)</td>
</tr>
<tr>
<td>$\alpha_4$</td>
<td>-0.0344</td>
<td>0.0143</td>
<td>(-0.0634, -0.0019)</td>
</tr>
<tr>
<td>$\alpha_5$</td>
<td>-0.0214</td>
<td>0.0178</td>
<td>(-0.0564, 0.0171)</td>
</tr>
<tr>
<td>$\alpha_6$</td>
<td>-0.0136</td>
<td>0.0164</td>
<td>(-0.0443, 0.0206)</td>
</tr>
<tr>
<td>$\alpha_7$</td>
<td>0.0060</td>
<td>0.0177</td>
<td>(-0.0261, 0.0430)</td>
</tr>
<tr>
<td>$\alpha_8$</td>
<td>-0.0310</td>
<td>0.0188</td>
<td>(-0.0639, 0.0067)</td>
</tr>
<tr>
<td>$\alpha_9$</td>
<td>-0.0169</td>
<td>0.0147</td>
<td>(-0.0450, 0.0111)</td>
</tr>
<tr>
<td>$\alpha_{10}$</td>
<td>-0.0152</td>
<td>0.0227</td>
<td>(-0.0559, 0.0304)</td>
</tr>
<tr>
<td>$\alpha_{11}$</td>
<td>-0.0039</td>
<td>0.0148</td>
<td>(-0.0321, 0.0268)</td>
</tr>
<tr>
<td>$\alpha_{12}$</td>
<td>-0.0150</td>
<td>0.0137</td>
<td>(-0.0432, 0.0137)</td>
</tr>
<tr>
<td>$\alpha_{13}$</td>
<td>-0.0243</td>
<td>0.0211</td>
<td>(-0.0669, 0.0149)</td>
</tr>
<tr>
<td>$\alpha_{14}$</td>
<td>-0.0138</td>
<td>0.0135</td>
<td>(-0.0412, 0.0127)</td>
</tr>
<tr>
<td>$\alpha_{15}$</td>
<td>-0.0457</td>
<td>0.0148</td>
<td>(-0.0734, -0.0162)</td>
</tr>
<tr>
<td>$\alpha_{16}$</td>
<td>-0.0403</td>
<td>0.0122</td>
<td>(-0.0650, -0.0166)</td>
</tr>
<tr>
<td>$\alpha_{17}$</td>
<td>-0.0574</td>
<td>0.0164</td>
<td>(-0.0857, -0.0232)</td>
</tr>
<tr>
<td>$\alpha_{18}$</td>
<td>-0.0368</td>
<td>0.0133</td>
<td>(-0.0641, -0.0087)</td>
</tr>
<tr>
<td>$\alpha_{19}$</td>
<td>-0.0277</td>
<td>0.0161</td>
<td>(-0.0582, 0.0057)</td>
</tr>
<tr>
<td>$\alpha_{20}$</td>
<td>-0.0474</td>
<td>0.0180</td>
<td>(-0.0855, -0.0123)</td>
</tr>
<tr>
<td>$\alpha_{21}$</td>
<td>-0.0459</td>
<td>0.0161</td>
<td>(-0.0767, -0.0137)</td>
</tr>
<tr>
<td>$\alpha_{22}$</td>
<td>-0.0233</td>
<td>0.0170</td>
<td>(-0.0550, 0.0102)</td>
</tr>
<tr>
<td>$\alpha_{23}$</td>
<td>-0.0051</td>
<td>0.0173</td>
<td>(-0.0406, 0.0280)</td>
</tr>
<tr>
<td>$\alpha_{24}$</td>
<td>-0.0581</td>
<td>0.0155</td>
<td>(-0.0921, -0.0314)</td>
</tr>
<tr>
<td>$\alpha_{25}$</td>
<td>-0.0075</td>
<td>0.0316</td>
<td>(-0.0701, 0.0488)</td>
</tr>
<tr>
<td>$\alpha_{26}$</td>
<td>0.1006</td>
<td>0.0648</td>
<td>(-0.0058, 0.2562)</td>
</tr>
<tr>
<td>$\alpha_{27}$</td>
<td>0.0989</td>
<td>0.0692</td>
<td>(-0.0058, 0.2844)</td>
</tr>
<tr>
<td>$\alpha_{28}$</td>
<td>0.0336</td>
<td>0.0432</td>
<td>(-0.0423, 0.1257)</td>
</tr>
<tr>
<td>$\alpha_{29}$</td>
<td>-0.0272</td>
<td>0.0246</td>
<td>(-0.0715, 0.0292)</td>
</tr>
<tr>
<td>$\alpha_{30}$</td>
<td>-0.0153</td>
<td>0.0341</td>
<td>(-0.0782, 0.0601)</td>
</tr>
<tr>
<td>$\alpha_{31}$</td>
<td>-0.0174</td>
<td>0.0265</td>
<td>(-0.0655, 0.0383)</td>
</tr>
<tr>
<td>$\alpha_{32}$</td>
<td>-0.0295</td>
<td>0.0184</td>
<td>(-0.0633, 0.0070)</td>
</tr>
<tr>
<td>$\sigma_\theta$</td>
<td>0.0806</td>
<td>0.0041</td>
<td>(0.0732, 0.0888)</td>
</tr>
<tr>
<td>$\bar{\phi}$</td>
<td>5.6779</td>
<td>0.3692</td>
<td>(5.0653, 6.4776)</td>
</tr>
<tr>
<td>$\sigma_\phi$</td>
<td>1.6243</td>
<td>0.2380</td>
<td>(1.2419, 2.1373)</td>
</tr>
</tbody>
</table>

Table 3.1. Parameter Estimates
CHAPTER 4

A STATE-SPACE MODEL OF PURCHASE TIMING FOR DIRECT MARKETING

4.1. Introduction

Marketing models of purchase timing often use inter-purchase times to indicate customer attractiveness. Customers who interact frequently with a firm are viewed as being more satisfied with their transactions and represent a more important stream of income than those who interact infrequently. In addition, customers who have made recent purchases are viewed as active, while those who have not placed orders for some time are viewed as being at risk of becoming inactive (Stone 1995; Miglautsch 2002). The recency and frequency of customer interaction are important elements of scoring algorithms in direct marketing used to rank customers in terms of their potential profitability (Hughes 1994; Schmittlein and Peterson 1994; Schijns and Schroder 1996).

The use of inter-purchase time as a dependent variable in models of purchasing timing offers a flexible approach to identifying profitable customers, but does so at a potential cost of inefficiently using the data. Statistical models of purchase timing typically summarize the average tendencies of customers and do not explore determinants beyond aggregate variables such as demographics. Recency and frequency data, however, can be influenced by effects related to consumption, satisfaction and competition, where
changes in any one of these elements influence the timing of a customer's purchase. If these elements where explicitly incorporated into a model of customer behavior, more precise estimates and predictions of customer behavior could occur.

A challenge in moving beyond summary measures of inter-purchase time in direct marketing is that information on factors such as consumption and satisfaction are not observed. It is difficult to build and estimate models of purchase timing that investigate its determinants because data in direct marketing only reveals a customer's interaction with a specific firm, and not all of the firms with offerings in a product category (Morisson and Schmittlein 1988; Jen and Wang 1998). It is possible, however, to conceive of latent processes that are thought to affect purchase timing decisions, and to investigate models where aspects of purchase timing are treated as the outcome of a data-generating process that has substantive meaning.

In this paper a model where time is used as an explanatory variable for purchases is introduced. That is, time is used as an independent variable and not as a dependent variable. The model is state-space in form, where the observation equation is a discrete binary indicator of customer purchase. The state variable is customer inventory, and purchases occur when the level of inventory crosses a re-order point. This state-space formulation is compared to a traditional model where inter-purchase times are the dependent variable. In both models, purchase quantities are used as explanatory variables, and heterogeneity is introduced as random-effects and estimated within a hierarchical Bayes model.

An important aspect of the proposed model is that inventory is consumed stochastically, not deterministically, over time. This randomness creates auto-correlation
in the state variable within each purchase cycle that allows us to obtain more precise estimates and improved forecasts of when purchases will occur. Autocorrelation is not present in traditional statistical models that view the inter-purchase time as the dependent variable, and treat a customer's history of inter-purchase times as a set of independent observations (Jain and Vilcassim 1991; Allenby, Leone and Jen 1999; Seetharaman and Chintagunta 2003).

The remainder of the paper is organized as follows. The proposed model is introduced in section 4.2. The information used, or extracted, from customer purchase records relative to a standard model where time is treated as a dependent variable are also discussed. Section 4.3 describes two datasets used to investigate performance of the models – a business-to-business dataset where purchases follow a relatively stable process, and a business-to-consumer dataset where purchase timing is more volatile. Section 4.3 also provides specific parameterization of the models and describes hierarchical Bayes estimation algorithms. Parameter estimates and predictive performance of the models are reported in section 4.4, and section 4.5 offers a discussion and some concluding comments.

4.2. Model Development

Figure 4.1 illustrates two ways of viewing purchase timing data. The traditional way is to view the time between purchases as the dependent variable, and to build models of purchase timing that explain inter-purchase times (Gupta 1991; Jain and Vilcassim 1991; Allenby, Leone and Jen 1999; Seetharaman and Chintagunta 2003; Vakratsas and
Let $r_{i,j}$ be the inter-purchase time between the $i^{th}$ purchase and the $(i+1)^{th}$ purchase by customer $j$ and let $w_{i,j}$ be the quantity purchased in the $i^{th}$ purchase. Then, a standard approach to modeling this data is to assume a distribution for the inter-purchase times.

$$r_{i,j} \sim \pi(t \mid \theta_j, w_{i,j})$$

(4.1)

where $\pi$ is a distribution with support on the positive real numbers, and $\theta_j$ are associated parameters for the $j^{th}$ customer. The use of $w_{i,j}$ as a conditioning argument in the distribution follows from the notion that inter-purchase times may vary depending on the previous quantity purchased.

An alternative approach that treats time as an independent variable involves viewing both purchase events and non-purchase events as informative relative to a latent process. Let $s_{t,j}$ be the level of the $j^{th}$ customer's inventory at time $t$, let $w_{t,j}$ be the amount purchased at time period $t$, and assume that a purchase occurs when the inventory level is below a threshold value. Assuming that inventory is depleted over time at a rate $\theta_j$, then a state-space model of purchase timing is:

**Observation Equation:**

$$y_{t,j} = \begin{cases} 1 & \text{if } s_{t,j} < 0 \\ 0 & \text{otherwise} \end{cases}$$

(4.2)

**State Equation:**

$$s_{t,j} = s_{t-1,j} - \theta_j + w_{t-1,j} y_{t-1,j} + \epsilon_{t,j}$$

(4.3)

where $t$ indicates the purchase timing of a household $j$ from time period 2 to time period $T_j$. The initial condition of state variable $s_{1,j}$ is assumed to be zero. $\theta_j$, the consumption rate, is expected to be positive. The threshold in the observation equation is arbitrarily set to zero because it is not identified. The inventory level declines by the value $\theta_j$ each
period, and increases due to previous purchases, \(y_{t-1,j}\). The presence of an error term, \(\varepsilon_{i,j}\), in the state equation indicates that the inventory level changes stochastically. Note that \(w_{i,j}\) in the interpurchase time model is the same as \(w_{t,j}\) in the state-space model (Harrison and Stevens 1976; West and Harrison 1997). Different subscripts, \(t\) and \(i\), are used to emphasize that the purchase amounts are incorporated into the models differently.

Both the standard inter-purchase time model (Equation (4.1)) and the state-space model (Equations (4.2) and (4.3)) can be used to forecast a customer's inter-purchase time. For the standard model, an inter-purchase time forecast is obtained directly from the assumed distribution \(\pi(r|\theta_j,w_{i,j})\):

\[
\Pr(r_{i^*,j}) = \int_{r-1}^{r} \pi(r | \theta_j, w_{i^*,j}) \, dr
\]

(4.4)

where \(w_{i^*,j}\) is the most recent quantity purchased, and the integration from \(r-1\) to \(r\) corresponds to units of time used in the analysis. In most cases, time is measured in weeks, and Equation (4.4) provides the probability that a purchase occurs in the week \(r\).

Inter-purchase times are represented in the state-space model as a series of non-purchases \((s > 0)\) followed by a purchase \((s < 0)\):

\[
\Pr(r_{i^*,j}) = \Pr(s_{i,j} = \cdots = s_{t-1,j} = 0, s_{t,j} = 1 | \theta_j, w_{i^*,j})
\]

(5)

where \(w_{i^*,j}\) is the most recent quantity purchased. Thus, the two models treat the data differently. The state-space model views both purchases and non-purchases as being informative about a customer's level of inventory and the associated timing of purchases. The inter-purchase time model uses an aggregated dependent variable defined entirely by the times associated with purchase events and not the non-purchases.
Since the data are used differently in the two models, the extent to which the state-space model leads to improved forecasts of inter-purchase times depends on the legitimacy of the state equation. If the latent inventory model provides an accurate description of what triggers purchases, the state-space model should lead to improved forecasts because it makes use of more of the data. However, if the latent inventory model is inaccurate, then the standard model should provide better forecasts. Performance of the models in datasets with different characteristics is investigated to determine the robustness of the latent variable inventory model.

4.3. Data and Model Specification

Two datasets are used to investigate properties of the model. The first dataset is from an office supply company engaged in business-to-business selling in the United States. A random sample of 651 firms was selected for analysis. The data span a period of two years, and the last purchase interval for each customer was used for predictive testing of the models. The average purchase frequency is 4.99 per year with a standard deviation of 3.96, and the average purchase amount is $99.16 with a standard deviation of $115.16.

The second dataset is from a direct marketing company specializing in cosmetics, shampoo, toothpaste and food supplements selling in Taiwan. A random sample of 606 customers was selected for analysis. The average purchase frequency is 15.32 per year with a standard deviation of 6.49, and the average purchase amount is $78.25 (Taiwan dollar) with a standard deviation of $13131.12 (Taiwan dollar). Thus, the two datasets
provide different environments with which to gage the robustness of the latent variable model. It is more likely, a-priori, that an inventory model applies to the first dataset and less likely that it applies to the second dataset.

4.3.1. State-Space Model Specification

Equations (4.2) and (4.3) are used as the basis for the state-space model. The initial inventory value for each respondent was set to zero, i.e., \( s_{1,j} = 0 \), and the parameter used to reflect the respondent consumption rate, \( \theta_j \), is constrained to be positive through an exponential transformation \( \theta_j = \exp(\theta^*_j) \) where \( \theta^*_j \) is estimated unconstrained.

Heterogeneity is introduced in all model parameters, including \( \sigma^2_j = \text{Var}(\epsilon_{t,j}) \):

\[
\begin{align*}
\theta^*_j &\sim \text{Normal}(\bar{\theta}, \sigma^2_\theta) \\
\bar{\theta} &\sim \text{Normal}(a, b^2) \\
\sigma^2_\theta &\sim \text{Inverted Gamma}(f_\theta, g_\theta) \\
\sigma^2_j &\sim \text{Inverted Gamma}(\kappa, \delta) \\
\kappa &\sim \text{Uniform}(3, \infty) \\
\delta &\sim \text{Inverted Gamma}(d_\delta, D_\delta)
\end{align*}
\] (4.6)

Thus, the model allows for heterogeneity in both the rate of consumption (\( \theta \)) and the regularity of consumption (\( \sigma^2_j \)). Parameters of the prior distributions and hierarchical Bayes algorithms for estimating the model are provided in the appendix.

A potential advantage of the state-space model is that each time period provides information about the latent state variable, \( s_{t,j} \). When a purchase occurs it is assumed that \( s_{t,j} < 0 \), and when a purchase does not occur \( s_{t,j} \geq 0 \). Thus, the distribution of the latent state variable, conditional on other model parameters, is multivariate normal:
\[ s_j \sim \text{Truncated Normal}(\mu_j, \Sigma_j) \]  

(4.7)

where

\[
s_j = \left( s_{1,j}, s_{2,j}, \cdots, s_{r,j} \right)^T
\]

\[
\mu_j = \begin{bmatrix}
-\theta_j + w_{i,j} y_{i,j} \\
-\theta_j + (w_{i,j} y_{i,j} + w_{2,j} y_{2,j}) \\
\vdots \\
-(T_j - 1)\theta_j + \sum_{i=1}^{T_j-1} w_{i,j} y_{i,j}
\end{bmatrix}
\quad \text{and} \quad
\Sigma_j = \begin{bmatrix}
1 & 1 & \cdots & 1 \\
1 & 2 & \cdots & 2 \\
\vdots & \vdots & \ddots & \vdots \\
1 & 2 & \cdots & T_j - 1
\end{bmatrix}
\]

with a correlated error structure. Dependence among the elements of the state variable results from inventory carryover from period to period.

An advantage of the state-space model estimated by Bayesian methods is that posterior estimates of state variable are readily available as a by-product of the estimation procedure. Thus, it is possible to make inferences about the level of customer inventory at any point in time and use this information to tailor customer-specific promotions.

4.3.2. Inter-Purchase Time Model Specification

Equation (4.1) is used as the basis for the inter-purchase time model, where a log-normal distribution for inter-purchase times is specified for \( \pi \):

\[
r_{i,j} = \ln(r_{i,j}^*) = \beta_{0,j} + \beta_{i,j} w_{i,j} + \varepsilon_{i,j}, \quad \varepsilon_{i,j} \sim N(0, \sigma_j^2)
\]

\[
r_{i,j}^* \sim LN(\beta_{0,j} + \beta_{i,j} w_{i,j}, \sigma_j^2)
\]

(4.8)

Heterogeneity is introduced into the regression coefficients, \( \beta_j \), and error variance, \( \sigma_j^2 \):
\[ \beta_j = (\beta_{0,j}, \beta_{1,j})' \sim \text{Normal}(\bar{\beta}, V_\beta) \]
\[ \bar{\beta} = (\bar{\beta}_0, \bar{\beta}_1)' \sim \text{Normal}(a, b^2) \]
\[ V_\beta \sim \text{Inverted Wishart}(g, H) \]  
\[ \sigma_j^2 \sim \text{Inverted Gamma}(\kappa, \delta) \]
\[ \kappa \sim \text{Inverted Gamma}(3, \infty) \]
\[ \delta \sim \text{Inverted Gamma}(d_\alpha, D_\alpha) \]  

Parameters of the prior distributions and hierarchical Bayes algorithms for estimating the inter-purchase time model are also provided in Appendix B.

### 4.4. Parameter Estimates and Predictive Results

Table 4.1 reports parameter estimates for the state-space and inter-purchase time models. Reported is the posterior mean and standard deviation of hyper-parameters that describe variation of the individual-level parameters. Graphical displays of the corresponding distributions are provided in Figures 4.2 and 4.3. The top panel of Figure 4.2 displays the distribution of the state-space consumption rate, \[ \theta_j^* = \ln \theta_j \], for both datasets, and the bottom panel displays the distribution of variance of the consumption rate, \[ \sigma_j^2 \]. The top two panels of Figure 4.3 displays the distribution of the regression coefficients, \[ \beta_{0j} \] and \[ \beta_{1j} \], that reflect the dependence of the inter-purchase time on the lagged purchase amount, \[ w_{i,j} \]. The bottom panel displays the distribution of the error term variance, \[ \sigma_j^2 \].

The estimates indicate that the purchase behavior of respondents the business-to-business dataset is much more regular than in the business-to-consumer dataset. For the
state-space model, heterogeneity in the log consumption rate ($\theta^*$) is an order of magnitude larger in the business-to-consumer dataset, with more than half the posterior estimates of the individual-level rates having negative algebraic sign. This indicates that the inventory model associated with the state-space model is more plausible in the business-to-business data than in the business-to-consumer data because the inventory depletion parameter, $\theta_j$, is expected to be positive, which corresponds to $\theta^*_j > 1$. The distribution of the variance parameters, $\sigma^2_j$, is approximately equal across the two datasets.

The heterogeneity distribution for the inter-purchase time model in Figure 4.3 also reveals a wider distribution of heterogeneity in the business-to-consumer data, with a large portion of the mass of the distributions for the regression coefficients located in the negative region. The distribution of the variance parameters is approximately equal in the two datasets, similar to that found in the state-space model.

One observation from each respondent’s purchase history was reserved for predictive testing of the models. Inter-purchase time forecasts from the two models were obtained using Equations (4.4) and (4.5) in conjunction with the corresponding posterior distribution of each model’s parameters. For both models, the most recent purchase quantity, $w_{i,j}$, was used constructing the predictive distribution of inter-purchase times. Figure 4.4 provides a comparison of the inter-purchase time forecasts of the two models. The horizontal axis of the graphs is the actual inter-purchase time, and the vertical axis is the difference in the predicted probabilities associated with these times. Points above the horizontal line in the plots indicate that the state-space model had higher predicted
probability, and points below the line are associated with the inter-purchase time model having higher probability.

Overall, the state-space model has greater predictive accuracy than the inter-purchase time model, particularly in the more stable, business-to-business dataset. The average forecast probability in this dataset for the state-space model is 0.0712 (std=0.21) versus 0.0437 (std=0.041) for the inter-purchase time model. The average forecast probability in the business-to-consumer dataset is 0.104 (std=0.148) for the state-space model and 0.145 (std=0.116) for the inter-purchase time model. The results displayed in Figure 4.4 indicate that the state-space model is particularly sensitive to respondents with short inter-purchase times. Modeling these respondents with a parametric distribution is problematic because these short inter-purchase times are in the tails of the distribution of the inter-purchase time model. The inter-purchase time model performs better in the moderate range of inter-purchase times (i.e., 3-10 weeks), and both models produce equivalent results for inter-purchase times greater than 10 weeks.

4.5. Discussion

Our results indicate that the state-space model can accurately describe customer behavior when the specification of the state equation is plausible for the data. Purchases in the business-to-business dataset are more regular than those in the business-to-consumer dataset, most likely because competitive effects on purchases are less frequent (e.g., vendor decisions are made for longer periods of time). When these competitive and other factors that influence purchase timing play less of a role, it is more likely that the
proposed latent inventory model offers a plausible description of respondent behavior (Aspinall, Nancarrow, and Stone 2001).

The advantage of using a latent variable model is that it provides more information about the consumer than standard inter-purchase time models. At any point in time, the latent inventory of a respondent can be predicted and used to tailor promotional offers (Plakoyiannaki and Tzokas 2002). Such predictions are not available from inter-purchase time models without additional assumptions concerning consumption rates and initial inventories, both of which are elements of the state-space model.

Latent variable models are an important aspect of consumer research whenever the determinants of behavior are an important aspect of study. To date, researchers have tended not to employ latent variable models with auto-correlated state variables because of the complexity involved in estimated models with a large "state-space." Bayesian methods of estimation, and in particular the method of data augmentation used in this paper, allow for more flexible model formulations that previously possible, hopefully leading to models with richer specification in the future.
Figure 4.1 Comparison between standard inter-purchase time model and state-space model: (a) Standard inter-purchase time model; (b) State-space model
Figure 4.1 Continued

Latent inventory level ($s_t$)

- $s_t > 0$
- $s_t = 0$
- $s_t < 0$
Figure 4.2 Heterogeneity distribution of state-space model parameters
Figure 4.3 Heterogeneity distribution of inter-purchase time model parameters
Figure 4.4 Comparison of model forecasts

(a) Business-to-Business Data

(b) Business-to-Consumer Data
<table>
<thead>
<tr>
<th>State-Space Model</th>
<th>Business-to-Business Data</th>
<th>Business-to-Consumer Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\bar{\theta}$</td>
<td>1.8708 (0.0927)</td>
<td>-8.45 (0.98)</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>0.2920 (0.0607)</td>
<td>146.16 (18.23)</td>
</tr>
<tr>
<td>$\kappa$</td>
<td>3.0396 (0.2799)</td>
<td>3.0001 (.0001)</td>
</tr>
<tr>
<td>$\delta$</td>
<td>0.0071 (0.0005)</td>
<td>0.0055 (.0001)</td>
</tr>
</tbody>
</table>

| Inter-Purchase Time Model | | |
|---------------------------|---------------------------|
| $\bar{\beta}_0$          | 1.76780 (0.03647)         | 0.8143 (0.0340)           |
| $\bar{\beta}_1$          | 0.05020 (0.0091)          | 0.0876 (0.0122)           |
| $\sigma_{11}$             | 0.05388 (0.0127)          | 0.4478 (0.0515)           |
| $\sigma_{12}$             | 0.00880 (0.0011)          | -0.1008 (0.0171)          |
| $\sigma_{22}$             | 0.00263 (0.0005)          | 0.0339 (0.0061)           |
| $\kappa$                  | 13.39592 (0.6366)         | 5.4334 (0.5486)           |
| $\delta$                  | 0.08368 (0.0042)          | 0.4778 (0.0564)           |

Table 4.1. Parameter Estimates (Posterior Standard Deviations)
CHAPTER 5

ESSAY 3: MODELING MEDIA INTERACTIONS AND PREFERENCE CHANGE IN PANEL DATA

5.1. Introduction

The communication of advertising media can be effective if a consumer’s response to media is known to marketing researchers. There are two challenges related to studying media effects on a consumer’s brand preferences and sensitivities to merchandising variables. The first challenge comes from the fact that a consumer’s purchase decision is affected by various information sources, such as TV commercials, printed advertisements, and the recommendation from friends. Researchers need a method to evaluate the interactions between each medium and marketing control variables such as price and display.

The second challenge is that, in most marketing applications, media effects cannot be directly measured by marketing researchers. This is because consumer preferences are latent and are affected by various media. For example, a consumer’s decisions reveal a ranking of alternatives in a decision task, and latent utilities of alternatives are derived from consumer preferences that are affected by multiple information sources. Researchers need to have a hierarchical model to study the relationship among consumer decisions, latent behavior processes, and media effects.
Practitioners usually study media interactions with merchandising variables in a pretest-posttest experiment (i.e., the experiment to measure the effectiveness of test commercials) where the measurements of dependent variables are collected from respondents before and after giving treatments. Treatment effects are evaluated by subtracting pretest measurements from posttest measurements. This approach can be easily implemented when measurements (e.g., blood pressure) can be collected directly in the interval or ratio scale. This approach is not appropriate for some marketing applications in which measurements (e.g., price sensitivity) are covered by observed brand choices and cannot be collected from consumers directly. Although treatment effects can be studied from analyzing the estimates of pre-post measurements, this approach cannot be used to study treatment effects on each respondent’s preferences and sensitivities to merchandising variables. Therefore, a new model is needed to study treatment effects for marketing problems.

The first research objective of this paper is to provide a method to study media interactions when pretest and posttest measurements are available to researchers. Media interactions are defined as the interactions between media and merchandising variables (e.g., price, display). The purpose of studying media interactions is to understand how media affect a consumer’s brand preferences and sensitivities to merchandising variables. Since the pre-post data are a pair of observations of a respondent, media interaction is studied by relating the pre-post measurements to the same random-effect realizations (e.g., pretest and posttest measurements are functions of same draws of random-effect distribution) in a latent variable model.
Two specifications of consumer heterogeneity are proposed to study media interactions with merchandising variables in this paper. The first approach is to specify the relationship between pretest and posttest measurements in an additive form. In this case, posttest measurements are equal to pretest measurements plus media effects. The second approach is to study media interactions by specifying a multiplicative relationship between pretest and posttest measurements. That is, posttest measurements are equal to pretest measurements multiplied by media effects. Multiplicative and additive relationships between pretest and posttest measurements have different implications to analyze media effects. When multiplicative relationship is specified, interaction effects between media and merchandising variables are a function of pretest measurements and media effects. However, when an additive relationship is specified, interaction effects between media and merchandising variables are not a function of pretest measurements. Therefore, multiplicative and additive specifications proposed in this paper can provide an approach to understand how a consumer’s brand preferences and sensitivities to merchandising variables are updated and carried over from a pretest study to a posttest study.

In addition to media interactions, this paper provides a solution to model “no-choice decisions” that are often observed in many real-world situations. The most common approach of dealing with no-choice decisions in literature is to exclude shopping trips that consumers do not purchase any offering from the product category of interest (e.g., Allenby and Lenk 1994). Discarding no-choice observations implies that no-choice observations do not provide information to consumer preference. However, no-choice decisions are also an outcome of decision-making process. Consumers decide not to
choose any item among alternatives for different reasons, such as the preferred alternative is not available in the choice set, the difficulty of making decision among the presented alternatives, the uncertainty to the best option, or the tendency to delay choice (Dhar and Simonson 2003). Removing no-choice observations can lead to an inaccurate prediction of choice probability because no-choice decisions reveal that consumers are better off by spending money on the outside good.

The alternative approach of handling no-choice decisions is to model no-choice decisions as a utility maximizing outcome in a multinomial model. This standard approach can be easily implemented if no-choice decisions are considered as an item competing with offerings (e.g., brand or brand-pack combinations) in a product category. However, a multinomial model cannot be used to study the tradeoff between the product class and the outside goods in the expenditure decision because this approach confounds the expenditure decision and the utility for brand. The expenditure decision is particularly important when the quantity decision is studied in a model. For example, the budget constraint, not the low utility of large quantity, may make a consumer only purchase a particular quantity.

No-choice decisions can also be interpreted as the consumer decisions of purchasing zero quantity of items in the product category. Several demand models have been proposed to deal with a consumer’s quantity decision in literature. Examples include Pudney (1989), Kim et. al. (2002), and Allenby et. al. (2003). However, these models proposed by Pudney (1989) and Kim et. al. (2002) cannot be used to accommodate no-choice decisions when a consumer’s total expenditure in a shopping trip is unknown. The demand model of Allenby et. al. (2003) does not have solution for no-choice decisions.
The second research objective of this paper is to develop a demand model to deal with consumer decisions of brand-pack and no-choice at the level of stock-keeping unit (SKU). A consumer’s quantity decision is discrete because the demand for brands is defined by a finite number of brand-pack combinations (e.g. 12 o.z. of brand A). When brand-pack choice is observed, the positive quantity of a brand is chosen by a consumer. When no-choice decision is observed, consumers purchase zero quantity in the product category of interest.

The proposed model is derived by assuming that no-choice decisions are outcomes of utility maximization and consumers allocate all expenditures on the outside goods if no-choice decisions are observed. A translated utility specification (Wales and Woodland 1983, Kim et. al. 2003) is employed to accommodate no-choice decisions. The likelihood function for the proposed model is derived from the assumption of normal random errors of the marginal utility. Error augmentation, a variant of data augmentation, is needed to deal with the estimation complexities associated with no-choice decisions and shared realizations of consumer heterogeneity that is used to study media interactions in a pre-post study.

Data provided by a leading company of consumer packaged goods are analyzed with the proposed model. This company would like to compete with its major competitor in a product category by introducing a new item that shares the same trademark with an established brand and provides similar product attributes as its major competitor. This marketing action is known as line extension. To predict the performance of new item and to evaluate the decision of line extension, data are collected from an information acceleration study. Information acceleration (IA) is a measurement methodology
proposed by Urban, Weinberg, and Hauser (1996) to forecast consumer reactions to a new product in a multimedia virtual-buying environment.

The IA study conducted by this company consists of three steps: pretest, information acceleration, and posttest. First, respondents are brought into a central site of the study. Respondents are asked to make choice among established brands in pretest virtual shopping trips. Then, different sources (e.g., TV, magazine) of product information of new product and established brands are simulated on computers in the step of information acceleration. Subject to an overall time limit, respondents can access information in any order, revisit information source, and spend as much time as they want on any information source. The content of information (e.g., TV commercial of brand b) is same for all respondents, but not all information sources are available to respondents. After the step of information acceleration, respondents are asked to make choice again among new product and established brands in posttest virtual shopping trips. Marketing merchandising variables such as price and display are varied among shopping trips. The choice decision of each respondent occurs at the level of stock keeping unit (SKU). Respondents are not forced to make choice because not all SKUs in the market are available in this study.

The demand model proposed in this paper is related to three issues in choice modeling. The first issue relates to the effect of multiple treatments for data collected in a pre-post study. Pre-post measurements are assumed to share the same random-effect realizations. By specifying the explicit relationship (e.g., additive) between pretest measurements and posttest measurements, the proposed model can be used to study the effect of multiple treatments on consumer preference change. The second issue relates to
line extension that is widely adopted in marketing practice. The proposed model provides a method to measure market response to a new product in an established product category. The last issue relates to the decision of brand-pack and no-choice for consumer packaged goods at the level of stock-keeping unit. A direct evaluation procedure and the translated Cobb-Douglas utility are proposed to accommodate no-choice decisions.

The reminder of the paper is organized as follows. Literature review of economic demand models and advertising effect models are in the next section. In section 3, the demand model for no-choice decisions is developed to study media interactions with merchandising variables. Section 4 describes data collected in an information acceleration study by a leading company of consumer packaged goods. The parameter estimates, the model comparison, the implication of data analysis are provided and discussed in section 5, and concluding remarks are offered in section 6.

5.2. Literature Review

This paper fits into the literature of modeling the effectiveness of advertising, preference change, the choice of discrete quantity, and no-choice decisions. In this section, the overview of related literature and the limitations are provided.

5.2.1. Preference Change and Consumer Heterogeneity

Preference change has been studied in the literature of dynamic models, the choice model of state dependence, and the structure of consumer heterogeneity. Dynamic models study the change of consumer behavior (e.g., brand switching) by specifying time-varying parameters. Examples of dynamic models include the dynamic change point
model for new product sales forecasting proposed by Fader, Hardie, and Huang (2004), and the dynamic model for utility evolution proposed by Liechty, Fong, and DeSarbo (2005). Research of dynamic models demonstrates that including the dynamic effect for preference change in a model can improve the performance of model (e.g., more accurate prediction). The drawback of dynamic models is that dynamic models cannot provide explanation to the causes of consumer preference change.

In addition to dynamic models, many choice models are developed to explain brand switch by studying preference change through the idea of state dependent. State dependent in choice behavior refers to the causality that the present purchase behavior depends on the past purchase behavior. Different behavioral explanations to state dependent have been proposed in literature. For example, state dependent may be caused by variety seeking behavior, habit persistence, or satiation (e.g., McAlister, 1982; Erdem 1996; Seetharaman, Ainslie, and Chintagunta 1999; Seetharaman 2004). Even though many behavioral explanations have been proposed in state dependent models, state dependent models cannot be used to study how consumer brand preference and sensitivity to marketing mix variables are impacted by marketing activities. It is because state dependent models assume that brand choice depends on lag brand choice and consumer preferences are invariant over time. However, if marketing activities, e.g., TV commercials, have impact on consumer preferences, the coefficients of state dependence and marketing variables should vary with marketing activities over time.

Consumer preferences are also studied from random-effect distribution. The most common example is to allow household-specific parameters depend on household-specific characteristics, such as demographic variables (e.g., Allenby and Ginter 1995;
Seetharaman, et. al. 1999). Instead of explaining preference variations among respondents using household-specific characteristics, Yang, Allenby, and Fennell (2002) investigate the influence of objective environments and motivating conditions on brand preferences by specifying the random-effect distribution with various relationships between respondents and objective environment. For example, the additive specification assumes that brand preferences equal to the summation of respondent effects and environmental effects, while the multiplicative specification assumes that coefficients of each respondent-environment combination is generated from a random-effect distribution with the location parameter defined by the multiplicative relationship between respondent effects and environment effects.

In this paper, the proposed approach of studying consumer preference change is close to the additive model proposed by Yang et. al. (2002) that assumes respondent effect is shared by various environments. Because pre-post data (e.g., brand choice) are a pair of observations of a respondent in a pre-post study, pre-post measurements are related by common random-effect realizations. Consumer preference change is studied by specifying relationships (e.g., additive or multiplicative) between pre-post measurements.

5.2.2. Advertising and Media Effects

The proposed model is also related to the literature of studying advertising and media effects. Three types of research have been conducted to study the effectiveness of advertising and media in marketing.

In the first type of research, sales response models are constructed to study the effect of market-level advertising expenditures on aggregate sales or market share.
Examples include Leone and Schultz (1980), Blattberg and Jeuland (1981), Assmus and Lehmann (1984), and Rao (1986). Sales response models are good at predicting the overall market response to advertising over time, but sales response models cannot be used to understand media effects at the level of individual households. It is because sales or market share is the aggregate outcome of consumer responses to marketing communications given advertising expenditures allocated to media. Consumers respond to media differently because consumers have different concerns and interests. The impacts of media on consumer choice is unknown if individual data (e.g., brand choice, media exposures) are not available to researchers. Therefore, sales response models are inappropriate for studying the impact of multiple media on consumer preference.

The second type of research focuses on the sales response to a single medium (e.g., Pedrick and Zufryden 1991) or to multiple media (e.g., Naik and Raman 2003). Pedrick and Zufryden (1991) develop an advertising response model to analyze the impact of advertising media plan and other marketing variables on the market performance of a brand. The model proposed by Pedrick and Zufryden (1991) integrates brand choice behavior, purchase incidence of a product category, and advertising exposures into a model.

However, the model developed by Pedrick and Zufryden (1991) cannot be used to study the impact of multiple media on consumer preferences for the following reasons. First, to study the impact of multiple media on consumer preferences, consumer preferences need to be a function of media exposures. But in Pedrick and Zufryden (1991), brand choices depend on media exposures. The impact of media on consumer preferences cannot be studied since consumer preferences do not vary with media.
Second, individual household advertising exposures are specified according to scheduled TV spots and the individual probability of viewing one of the scheduled spot. To study the impact of multiple media, the model of advertising exposures behavior proposed by Pedrick and Zufryden (1991) needs to accommodate multiple media and the probabilities of viewing media. The aggregate population exposures distribution specified with Beta-Binomial distribution cannot be extended to other media, such as printed ads, because not all media have planned schedules like TV commercials and the frequency for a consumer to receive product information of some medium is difficult to measure.

Naik and Raman (2003) study the synergy of multimedia by a dynamic sales response model. The synergy effect is evaluated by its contribution to the total sales of a brand. The limitation of the model proposed by Naik and Raman (2003) is that their model cannot be extended to analyze the synergy effect of more than two media because media-specific carryover effects are unidentifiable and inestimable. Besides, most non-Bayesian estimation methods may break down because the increase in the number of media results in overparameterization of the model (e.g., the number of variables is greater than the sample size).

The third type of research uses psychology theories to explain the effectiveness of advertising in the laboratory experiments. The research examines the advertising effect on the change of memory, information processing, knowledge structure, attitude strength, belief strength, and/or confidence of a consumer (e.g., Wright and Lynch 1995; Lee 2002; Chakravarti and Janiszewski 2004). However, even though the advertising effect is proven to be statistically significant in a laboratory experiment, the lab-experimental research cannot provide meaningful estimates of the impact of advertising on the demand
of a brand because the size of advertising effect cannot be estimated and the finding cannot be used to predict market response (e.g., the increase in demand).

The approach of studying media effects proposed in this paper provides a solution to model the impact of multiple media on a consumer’s brand preferences and sensitivities. Since the data are collected in a laboratorial pre-post study, media selected by a respondent are available to researchers. A consumer’s brand preference change is studied by relating pre-post measurements to the same random-effect realizations, and by specifying posttest measurements as a function of media exposures and pretest measurements. The proposed approach has two advantages. First, the effect of multiple media can be evaluated simultaneously. Second, household-level estimates of consumer heterogeneity can be obtained to facilitate the polity evaluation.

5.2.3. Discrete Quantity

In reality, consumers are not necessarily forced to make purchase decision among all available items. Consumers decide not to choose any item among alternatives for different reasons, such as the preferred item is not available in the choice set, the difficulty of making decision among the presented alternatives, the uncertainty to the best option, or the tendency to delay choice (Dhar and Simonson 2003). Moreover, in the product of consumer packaged goods, consumers are restricted to make choice among available brand-pack combinations, and only limited options of pack size (e.g., 12 oz, 6 packs) are provided by the manufactures. Therefore, in this paper, a demand model is developed to deal with consumer decision of brand-pack and no-choice in the product category of consumer packaged good.
Economics models have been developed for the case that the consumer behavior reveals a discrete switch between the choice of positive quantity and zero quantity for no choice (Pudney 1989). Most of demand models deal with the decision of purchasing positive quantity of a product, and very few models can handle no-choice decisions. The first approach is to solve quantity decisions with the standard Kuhn-Tucker conditions. Examples of models derived from the Kuhn-Tucker conditions include Wales and Woodland (1983) and Kim, et. al. (2002).

Assume that $B$ is the total number of brands in a product category, $x$ is the vector of quantity demanded with element $x_j$, $p$ is the vector of price of quantity demanded with element $p_j$, $u(x)$ is the utility function of observed demand $x$, $E$ is the total expenditure, the adding-up budget constraint is $p'x = E$. To solve the optimal demand by the Kuhn-Tucker conditions, the Lagrangian for the problem shown in Equation (5.1) needs to be formed

$$L = u(x) - \lambda (p'x - E)$$

(5.1)

Then the standard Kuhn-Tucker first-order conditions can be derived by differentiating the Lagrangian as follows:

$$\frac{\partial L}{\partial x_j} = u(x_j) - \lambda p_j < 0 \text{ if } x_j^* = 0$$

$$\frac{\partial L}{\partial x_j} = u(x_j) - \lambda p_j = 0 \text{ if } x_j^* > 0$$

(5.2)

where $x_j^*$ is the optimal demands for the $j^{th}$ good, and $\lambda$ is the Lagrangian multiplier. By introducing multiplicative normal error into marginal utility

$$\ln u(x_j) = \ln \bar{u}(x_j) + \varepsilon_j, \varepsilon_j \sim N(0,1)$$

(5.3)
, dividing Equation (5.2) by price, and taking log, the standard Kuhn-Tucker first-order conditions presented in Equation (5.2) can be rewritten as

\[
\ln \left( \frac{\bar{u}(x_j)}{p_j} \right) + \varepsilon_j < \ln \lambda \text{ if } x_j^* = 0 \\
\ln \left( \frac{\bar{u}(x_j)}{p_j} \right) + \varepsilon_j = \ln \lambda \text{ if } x_j^* > 0
\]  

(5.4)

Assume that the price of the outside goods is 1 and the outside goods are always purchased. Subtracting the Kuhn-Tucker conditions for the outside goods from the others, Equation (5.4) is equivalent to

\[
\nu_j < h_j(x^*, p) \text{ if } x_j^* = 0 \\
\nu_j = h_j(x^*, p) \text{ if } x_j^* > 0
\]  

(5.5)

where \( h(x^*, p) = \ln \left( \frac{\bar{u}(x_j)}{p_j} \right) - \ln (\bar{u}(x_{out})) \) and \( \nu_j = \epsilon_{out} - \epsilon_j \).

Assume that \( \epsilon_j \) and \( \epsilon_{out} \) are distributed with standard normal. The likelihood of demand \( x^* \) can be derived from the p.d.f. of \( \nu=(\nu_1, \nu_2, \ldots, \nu_B) \) and the Kuhn-Tucker conditions presented in Equation (5.5). However, the limitation of the Kuhn-Tucker conditions is that, when the total expenditure in a shopping trip is unknown, the first-order conditions shown in Equation (5.2) cannot be solved and cannot be used to derive the likelihood of quantity decisions.

An alternative approach is to study the demand of brand-pack by a direct evaluation procedure of Allenby et. al. (2003). Assume that \( x=(x_1, x_2, \ldots, x_B) \) is the vector of the purchase amount of each brand, the price of purchasing quantity \( x \) of brand \( b \) is \( p(x_b) \), \( B \) is the total number of brands in the study, \( w \) represents the amount of the outside good
purchased, and $T$ is the total expenditure. Given the “adding-up” budget 

constraint $\sum_{b=1}^{B} p(x_b) + w = T$ and log-marginal utility of brand $b$ is stochastic with an additive brand-specific error $\epsilon_b$, the log Cobb-Douglas utility function applied in Allenby et. al. (2003) is 

$$\ln u(x_b, w) = \ln u(x_b) + \alpha \ln u(w)$$ 

$$= \nu_b + \ln(x_b) + \alpha \ln(T - p(x_b)) + \epsilon_b$$ 

(5.6) 

Assume that $x^*_b$ is the optimal quantity of brand $b$. $x_{kb}$ is the quantity $x_k$ of brand $b$. Allenby et. al. (2003) suggests identifying the optimal quantity given that a brand is purchased by deterministically searching for the quantity that maximizes the following expression 

$$x^*_b = \arg \max_k \left( \nu_{kb} + \ln(x_{kb}) + \alpha_T \ln(T - p(x_{kb})) \right)$$ 

(5.7) 

The error $\epsilon_b$ cancels in Equation (5.7) because the error is assumed to be the same for all packsize with the same brand name. 

Assume that $\epsilon_b$~$\text{EV}(0,1)$. Conditional on the optimal quantity of each brand determined by Equation (5.7), the choice probability of the observed demand $x_i$ can be evaluated by the standard multinomial logit model 

$$\Pr(x_i) = \frac{\exp \left( \nu_i + \ln(x_i) + \alpha_T \ln(T - p(x_i)) \right)}{\sum_{b=1}^{B} \exp \left( \nu_b + \ln(x^*_b) + \alpha_T \ln(T - p(x^*_b)) \right)}$$ 

(5.8) 

Then the standard parameter estimation can be preceded given the optimal product quantity of each brand. 

The approach of Allenby et. al. (2003) provides a solution to analyze the data with discrete quantity decisions. However, the solution procedure and the likelihood of
Allenby et. al. (2003) cannot be used to analyze the data with no-choice decisions for two reasons. First, the Cobb-Douglas utility given in Equation (5.6) can solve for the positive utility-maximizing quantity for at least one brand, but it cannot solve the utility maximization for purchasing zero quantity of goods in the product category of interest. It is because zero quantity of goods cannot be directly evaluated by Equation (5.7). Therefore, a utility function is needed to allow the utility-maximizing solution for both brand-pack choices and no-choice decisions.

Second, the probability of observing the decision of zero quantity still cannot be derived in the demand model of Allenby et. al. (2003) even though a proper utility function that can deal with no-choice decisions is used. It is the case because, if the option of no choice is considered as an alternative competing with brands in the product category, an extra error needs to be added on the utility specification of the outside good. That is, there will be two errors, $\varepsilon_b$ and $\varepsilon_{out}$, present in Equation (5.6). As a result, the adding-up budget restriction is no longer satisfied. The likelihood of demand cannot be derived from the distribution assumption of random utility errors because substituting the outside good with the budget constraint cannot reduce the dimensionality of the model (see Kim et. al 2002 and Allenby et. al. 2003).

In this paper, a demand model is proposed to study consumer decision of brand-pack choice and no-choice. The proposed model neither relies on solving the standard Kuhn-Tucker conditions nor treats the option of no choice as one of the multinomial choice alternatives. Instead, the model is constructed according to the economic theorem of utility maximization. A translated utility specification (Wales and Woodland 1983, Kim et. al. 2003) is employed to accommodate no-choice decisions. The likelihood
function for this model is derived from normal random errors of the marginal utility. Error augmentation, a variant of data augmentation, is needed to deal with the estimation complexities associated with no-choice decisions and shared realization of consumer heterogeneity that is used to study media interactions in a pre-post study.

5.3. Model Development

A Cobb-Douglas utility function is used to model the tradeoff in expenditure between the offerings in the product category of interest and the outside good. A Cobb-Douglas utility function is defined as

$$ \ln u(x, w) = \alpha_1 \ln u(x) + \alpha_2 \ln u(w) $$

(5.9)

where $$\alpha_1 + \alpha_2 = 1$$, and $$x=(x_1, x_2, ..., x_B)$$ is the vector of the purchase amount of each brand, $$B$$ is the total number of brands in the study, and $$w$$ represents the amount of the outside good purchased. The optimal expenditure of a consumer on $$B$$ brands and the outside good $$w$$ are determined by solving the utility maximization of Equation (5.9) subject to the budget constraint

$$ \sum_{b=1}^{B} p(x_b) + w = T $$

(5.10)

where $$p(x_b)$$ is the price of $$x_b$$ unit of brand $$b$$, the price for the outside good is 1, and $$T$$ is the total expenditure. Let the linear sub-utility $$u(x)=\psi'x$$. According to the standard choice model, the log-marginal utility of brand $$b$$, $$\ln \psi_b$$, is assumed to be stochastic with an additive error. That is, $$\ln \psi_b = \nu_b + \epsilon_b$$. 

80
The likelihood for the observed demand is determined by the quantity decision of brands \((x_1, x_2, \ldots, x_B)\) in the product category and the outside good \((w)\). The distribution assumption of random utility error \(\varepsilon_b\) is used to derive the probability of observed choice. The budget constraint defined in Equation (5.10) allows the likelihood of utility maximization demand \((x,w)\) to be derived from \(B\) errors by substituting the outside good \(w\) in Equation (5.9) by \(T - \sum_{b=1}^{B} p(x_b)\). Assume that only one brand (brand \(b\)) is chosen, the log Cobb-Douglas utility function can be expressed as

\[
\ln u(x_b) = \alpha_1 \ln u(x_b) + \alpha_2 \ln \left(T - p(x_b)\right) \\
= \alpha_1 \ln (\psi_b x_b) + \alpha_2 \ln \left(T - p(x_b)\right) \\
= \alpha_1 \left(\ln \psi_b + \ln (x_b)\right) + \alpha_2 \ln \left(T - p(x_b)\right) \\
= \alpha_1 (\nu_b + \varepsilon_b) + \alpha_1 \ln (x_b) + \alpha_2 \ln \left(T - p(x_b)\right) + \varepsilon_b \\
= \alpha_1 \nu_b + \alpha_1 \ln (x_b) + \alpha_2 \ln \left(T - p(x_b)\right) + \varepsilon_b
\] (5.11)

Assume that the random utility error \(\varepsilon_b\) is distributed with standard normal to achieve empirical identification. Since the log-utility \(\ln u(x_b)\) for all brands in the study can be multiplied by an arbitrary constant without altering the order of utilities implied by the model, the exponent \(\alpha_1\) of the utility for the consumption of goods in the product category of study is set to 1 to impose the normalization restriction for the model.

Because the log Cobb-Douglas utility function in Equation (5.11) can only have the utility maximization solution for the positive purchase quantities of all brands, the translated log Cobb-Douglas utility function specified in Equation (5.12) is used to accommodate the decision of purchasing zero quantity by translating the indifference curve one unit toward the first quadrant (see Kim et. al. 2002).

\[
\ln u(x_b) = \alpha_1 \nu_b + \alpha_1 \ln (x_b + 1) + \alpha_2 \ln \left(T - p(x_b)\right) + \varepsilon_b
\] (5.12)
Let \( v_b = z_b \beta \). The matrix \( z_b \) contains the brand intercept, the merchandising variables, and the product attributes of brand \( b \). The translated log Cobb-Douglas utility function applied in the proposed model is

\[
\ln u(x_b) = z_b \alpha \beta + \alpha_1 \ln (x_b + 1) + \alpha_2 \ln \left( T - p(x_b) \right) + \epsilon_b \\
= z_b \beta_z + \alpha_1 \ln (x_b + 1) + \alpha_2 \ln \left( T - p(x_b) \right) + \epsilon_b, \quad \epsilon_b \sim N(0, 1) \tag{5.13}
\]

In Equation (5.13), \( \alpha / \beta \) is replaced by \( \beta_z \) since \( \alpha \) and \( \beta \) cannot be jointly identified.

Assume that the zero quantity is an option of quantity choice provided by a brand and that the random error \( \epsilon_b \) only affects brand choice, the solution procedure suggested by Allenby et. al. (2003) can be used to search the optimal quantity (include zero quantity) of each brand and then determine the bundle of brand-pack that provides the maximum utility. Let \( b \) denote the brand, \( x_{kb} \) refer to the quantity \( x_k \) of brand \( b \), and \( x^*_b \) represent the utility-maximizing quantity of brand \( b \). The utility maximizing quantity of brand \( b \) can be determined by searching over all available packsizes, including zero quantity, of brand \( b \):

\[
x^*_b = \arg \max_k \left\{ z_{kb} \beta_z + \alpha_1 \ln (x_{kb} + 1) + \alpha_2 \ln \left( T - p(x_{kb}) \right) + \epsilon_b \right\} \\
= \arg \max_k \left\{ z_{kb} \beta_z + \alpha_1 \ln (x_{kb} + 1) + \alpha_2 \ln \left( T - p(x_{kb}) \right) \right\} \tag{5.14}
\]

The error \( \epsilon_b \) cancels because the error is assumed to be the same for all packsizes with the same brand name.

In this paper, the likelihood of observed utility-maximizing demand of brand-pack choices and no-choice decisions is derived from the distribution assumption of random utility error \( \epsilon_b \). The brand-pack choice implies that the utility of observed choice gives the maximum utility among all available brand-pack combinations in the product category. The probability of observing the decision of purchasing \( x_k \) quantities of brand \( b \) is
\[
\Pr(x_{jb}) = \Pr\left(\ln u(x_{jb} = x_{j}^{*}b) > \max\left\{\ln u(x_{b}^{*}) \forall b' \neq b\right\}\right) \tag{5.15}
\]

The no-choice decision implies that consumers spend all expenditures on the outside good and do not purchase any item from the product category of study. Therefore, the no-choice decision is not associated with any brand, and it reveals a partial ranking of all brand-pack bundles. That is, the maximum utility of the offerings with zero quantity must be greater than the maximum utility of the offerings with positive quantity.

Let \(i\) denote the set of all brands with optimal quantity in the study, \(i'\) denote the set of brands with optimal quantity equal to zero, and \(i''\) denote the set of brands with positive optimal quantity. The set of brands with zero optimal quantity and the set of brands with positive optimal quantity are mutual exclusive. The probability of observing a no-choice decision is

\[
\Pr(0) = \Pr\left(\max_{j'} \left\{\ln u\left(x_{j'}^{*} = 0\right) \right\} > \max_{j} \left\{\ln u\left(x_{j}^{*} \neq 0\right) \right\}\right) \text{ and } \{i' \cup i''\} = \{i\} \tag{5.16}
\]

The model specified in Equation (5.13), (5.15) and (5.16) is extended to study treatment effects in a pre-post study. Let the superscript \(j=1\) represent the pretest study and the superscript \(j=2\) represent the posttest study respectively. Let the subscript \(t\) represent the \(t^{th}\) shopping trip, \(x_{bkt}\) be a choice of a positive product quantity of brand \(b\) at the shopping trip \(t\), and \(x_{bkt}^{*}\) be the optimal quantity of brand \(b\) at the shopping trip \(t\). The model proposed to study treatment effects for data collected in a pre-post study consists of two likelihoods. The likelihood of observing brand-pack choices and no-choice decisions in the study \(j\) are

\[
\Pr\left(x_{bkt}^{(j)}\right) = \Pr\left(\ln u\left(x_{bkt}^{(j)} = x_{bkt}^{*}\right) > \max\left\{\ln u\left(x_{bkt}^{*}\right) \forall b' \neq b\right\}\right)
\]

\[
\Pr(0) = \Pr\left(\max_{j'} \left\{\ln u\left(x_{j'}^{*} = 0\right) \right\} > \max_{j} \left\{\ln u\left(x_{j}^{*} \neq 0\right) \right\}\right) \text{ and } \{i' \cup i''\} = \{i\} \tag{5.17}
\]

---

83
and the translated log Cobb-Douglas utility function for the study $j$ is

$$\ln u^{(j)}(x_{ilst}) = z^{(j)} \beta^{(j)} + \alpha_s^{(j)} \ln(x_{ilst} + 1) + \alpha_t^{(j)} \ln\left(T^{(j)} - p\left(x_{ilst}^{(j)}\right)\right) + \epsilon_t^{(j)} + \epsilon_{ist}^{(j)} \sim N(0,1)$$

(5.18)

5.3.1. Treatment Effect

Since pretest and posttest data are a pair of observations of a respondent, treatment effects on post measurements are studied by relating pretest and posttest measurements to the common random-effect realizations and by specifying the relationship between pretest and posttest measurements with an explicit function. Multiplicative and additive relationships of pre-post measurements are examined in this paper. To reduce the notation burden, the following example is used to illustrate the difference between the multiplicative consumer heterogeneity and the additive consumer heterogeneity.

Assume that the pretest log linear utility function is

$$\ln u^{(1)} = z^{(1)} g^{(1)} + \epsilon^{(1)}$$

(5.19)

and the posttest log-utility linear function is

$$\ln u^{(2)} = z^{(2)} g^{(2)} + \epsilon^{(2)}$$

(5.20)

where $g^{(1)}$ is a pretest measurement and $g^{(2)}$ is a posttest measurement of a respondent, and the pretest independent variable $z^{(1)}$ and the posttest independent variable $z^{(2)}$ correspond to the same marketing mix variable (e.g., deal).

Let $m$ denote the treatment and $\gamma$ denote the treatment effect. The multiplicative consumer heterogeneity is given by

$$g^{(2)} = (1 + \gamma m) g^{(1)} = g^{(1)} + \gamma m g^{(1)}$$

(5.21)
An alternative way to study treatment effects is to specify the relationship between the pretest and posttest measurements using the additive form. That is,

$$\mathcal{G}^{(2)} = \mathcal{G}^{(1)} + \gamma m$$  \hspace{1cm} (5.22)

The additive specification of consumer heterogeneity is very close to the additive model proposed by Yang et. al. (2002). As shown in Equation (5.22), the residual of posttest random effect is the pretest parameter in the additive specification. However, the multiplicative consumer heterogeneity in Equation (5.21), the media effect is a function of the pretest measurement. In addition, when no treatment $m$ is received ($m=0$), the posttest measurement is equal to the pretest measurement in both Equation (5.21) and (5.22). It implies that a consumer’s preferences are not altered by treatments if no treatments are received.

If the multiplicative heterogeneity specified in Equation (5.21) is applied, the posttest log linear utility can be rewritten as

$$\ln u^{(2)} = z^{(2)} \mathcal{G}^{(2)} + \varepsilon^{(2)}$$

$$= z^{(2)} \mathcal{G}^{(1)} + z^{(2)} \gamma m \varepsilon^{(1)} + \varepsilon^{(2)}$$  \hspace{1cm} (5.23)

Alternatively, if the additive heterogeneity in Equation (5.22) is used to specify the relationship between pretest and posttest measurements, the posttest log linear utility function can be expressed as

$$\ln u^{(2)} = z^{(2)} \mathcal{G}^{(2)} + \varepsilon^{(2)}$$

$$= z^{(2)} (\mathcal{G}^{(1)} + \gamma m) + \varepsilon^{(2)}$$  \hspace{1cm} (5.24)

$$= z^{(2)} \mathcal{G}^{(1)} + z^{(2)} \gamma m + \varepsilon^{(2)}$$

Multiplicative and additive specifications of consumer heterogeneity have different implications of consumer preference changes. Equation (5.23) shows that, in the multiplicative specification, the interaction effect between the treatment $m$ and the
independent variable \( z^{(2)} \) is a function of the pretest measurement \( \vartheta^{(1)} \). However, when the additive specification is applied, Equation (5.24) shows that the interaction between the treatment \( m \) and the independent variable \( z^{(2)} \) is not influenced by the pretest measurement \( \vartheta^{(1)} \). Therefore, multiplicative and additive specifications of consumer heterogeneities provide an opportunity to study how consumer preference change \((\vartheta^{(1)}, \vartheta^{(2)})\) before and after giving treatments \( (m) \) and whether consumer preferences in a pretest study \((\vartheta^{(1)})\) have impacts on interactions between treatments \( (m) \) and posttest merchandising variables \((z^{(2)})\).

Let the \( b^{th} \) row vector of design matrix \( z^{(j)*}_{ht} \) contain the independent variables \( z^{(j)*}_{bht} \) for product attributes and merchandising variables, \( \ln(x^{(j)}_{bht}+1) \), and \( \ln(\mathcal{T}^{(j)}-p(x^{(j)}_{bht})) \) of brand \( b \) received by respondent \( h \) at the shopping trip \( t \) of study \( j \). The translated log Cobb-Douglas utility function for optimal quantities of all brands in the \( t^{th} \) shopping trip of the study \( j \) can be expressed in matrix form as

\[
\ln u(x^{(j)*}_{ht}) = z^{(j)*}_{ht} \beta^{(j)}_{ht} + \varepsilon^{(j)*}_{ht} \sim N(0, I^{(j)})
\]  

Let \( m_{bh} \) denote the treatments (e.g., TV commercials) of brand \( b \), and \( k_{zh} \) denote the treatments associated with merchandising variables \( z_{k} \). Given the multiplicative relationship between pretest and posttest measurements, the posttest measurements are specified as

\[
\beta^{(2)}_{h} = C_{h} \beta^{(1)}_{h}, \quad C_{h} = \text{diag} \left( 1 + \gamma_{h} m_{bh}, \ldots, 1 + \gamma_{h} m_{bh}, 1 + \theta_{1h} m_{z_{1h}}, \ldots, 1 + \theta_{\text{var},h} m_{z_{\text{var},h}}, 1 + \theta_{,h} m_{sh} \right)
\]  

where \( C_{h} \) is a diagonal matrix that contains media effects \( \gamma_{h} \) of respondent \( h \) for brand 1 to brand \( B \) and media effects \( \theta_{i,h} \) of respondent \( h \) for all independent variables and \( \ln(x^{(2)}_{bht}+1) \). Since the coefficient of \( \ln(\mathcal{T}^{(2)}_{h}-p(x^{(2)}_{bht})) \) needs to be positive to have a valid
Cobb-Douglas utility function, the coefficient corresponding to \( \ln(T^{(2)}_{h}\cdot p(x^{(2)}_{bh})) \) is specified as

\[
\beta^{(2)}_{T,h} = \exp \left( (1 + \theta^{(1)}_{T,h} m_{T,h}) \cdot \beta^{(1)*}_{T,h} \right)
\]
\[
\beta^{(1)}_{T,h} = \exp \left( \beta^{(1)*}_{T,h} \right)
\]

(5.27)

where the media effect \( \theta_{T,h} \) of respondent \( h \) for \( \ln(T^{(2)}_{h}\cdot p(x^{(2)}_{bh})) \), and \( \beta^{(1)}_{T,h} \) is the coefficient corresponding to \( \ln(T^{(1)}_{h}\cdot p(x^{(1)}_{bh})) \).

For the additive consumer heterogeneity, the relationship between pretest and posttest measurements is specified as:

\[
\beta^{(2)}_{h} = \beta^{(1)}_{h} + \Delta_{h}
\]
\[
\Delta_{h} = \left( \gamma_{h} m_{1h}, \cdots, \gamma_{h} m_{Bh}, \theta_{ih} m_{zh}, \cdots, \theta_{nvar,h} m_{zvarh}, \theta_{i,h} m_{x,h} \right)
\]

(5.28)

where \( \Delta_{h} \) is a column vector that contain media effects \( \gamma_{h} \) of respondent \( h \) for brand 1 to brand \( B \) and the media effect \( \theta_{ih} \) of respondent \( h \) for all independent variables and \( \ln(x_{bh}+1) \). Since the coefficient for \( \ln(T^{(2)}_{h}\cdot p(x^{(2)}_{bh})) \) needs to be positive to have a valid Cobb-Douglas utility function, the parameter corresponding to \( \ln(T^{(2)}_{h}\cdot p(x^{(2)}_{bh})) \) is specified as

\[
\beta^{(2)}_{T,h} = \exp \left( \beta^{(1)*}_{T,h} + \theta_{T,h} m_{T,h} \right)
\]
\[
\beta^{(1)}_{T,h} = \exp \left( \beta^{(1)*}_{T,h} \right)
\]

(5.29)

where \( \theta_{T,h} \) is the media effect of respondent \( h \) for \( \ln(T^{(2)}_{h}\cdot p(x^{(2)}_{bh})) \), and \( \beta^{(1)}_{T,h} \) is the coefficient corresponding to \( \ln(T^{(1)}_{h}\cdot p(x^{(1)}_{bh})) \).
5.3.2. Generating Latent Utility

The proposed model with multiplicative consumer heterogeneity (Equation (5.17), (5.18), (5.26), and (5.27)) and the proposed model with additive consumer heterogeneity (Equation (5.17), (5.18), (5.28), and (5.29)) are estimated by the Bayesian method of error augmentation and Markov Chain Monte Carlo method. The solution procedure of estimating the proposed models consists of three steps. The first step involves the deterministic search for the optimal product quantity for each brand. Different from Allenby et. al. (2003), the product quantity equal to zero is considered as a candidate of the optimal quantity of a brand. In the second step, the latent utilities are generated according to a multinomial probit model. Finally, the error augmentation and the proposed check-and-see algorithm are applied to estimate parameters.

The challenge of estimating the proposed models with no-choice observations is that a no-choice decision reveals a partial ranking of the alternatives in the optimal set identified by the deterministic search. In other words, a no-choice decision is observed because the maximum utility of brand-pack bundles with zero quantity is greater than the maximum utility of brand-pack bundles with positive quantity. Since pretest and posttest latent utilities can be generated by the same approach, superscripts used to denote the pretest and posttest data are dropped in this section to reduce notation burden.

When the positive quantity of choice is observed, the latent utility of a brand can be generated by using the standard approach. That is, after searching for the optimal product quantity of each brand, the latent utility of optimal product quantity of each brand can be generated from the univariate normal with truncation region specified in Equation (5.15).
Let $b'$ denote brand not purchased. The log latent utility for purchasing quantity $x_k$ of brand $b$ at the shopping trip $t$ can be generated from

$$\ln u(x_{kt}) \sim N\left(\ln \bar{u}(x_{kt}), 1\right) \cdot I\left(\ln u(x_{kt}) = x_{kt}^* \right) > \max \left\{ \ln u(x_{kt}^*), \forall b \neq b' \right\} \quad \text{(5.30)}$$

where

$$\ln \bar{u}(x_{kt} = x_{kt}^*) = \alpha \ln(x_{kt}^* + 1) + \beta \ln(T - p(x_{kt}^*))$$

Let $x_{kt}^*$ be the optimal quantity of brand $b'$ that is not purchased at the shopping trip $t$. The log latent utility for any other brand not purchased at the shopping trip $t$ can be generated from

$$\ln u(x_{kt}^*) \sim N\left(\ln \bar{u}(x_{kt}^*), 1\right) \cdot I\left(\ln u(x_{kt}^*) > \ln u(x_{kt}^*) \right) \quad \text{(5.31)}$$

where

$$\ln \bar{u}(x_{kt}^*) = \alpha \ln(x_{kt}^* + 1) + \beta \ln(T - p(x_{kt}^*))$$

The decision of no-choice implies that at least one brand has optimal product quantity equal to zero, and the maximum log utility of brands with zero optimal product quantity must be greater than the maximum log utility of brands with positive optimal product quantity. Let $i$ denote the set consisting of optimal quantities of all brands in the study, $i'$ denote the set of brands with optimal quantity equal to zero, and $i''$ denote the set of brands with positive optimal quantity. The set of brands with zero optimal quantity and the set of brands with positive optimal quantity are mutual exclusive. The difficulty of generating the latent utilities for no-choice decisions from the fact that many different permutations satisfy this relation defined in Equation (5.16):

$$\max_{i'} \left\{ \ln u(x_{r_i} = 0) \right\} > \max_{i''} \left\{ \ln u(x_{r_i}^* \neq 0) \right\} \quad \text{and} \quad \left\{ i' \cup i'' \right\} = \{ i \}$$

To generate the latent utilities when no-choice decisions are observed, the following conditions need to be considered after the optimal product quantity of each brand is determined in the first step of solution procedure.
**Condition 1:** If all brands have zero optimal product quantity, \( \ln u(x_{bt}^*) \) can be generated from

\[
\ln u(x_{bt}^* = 0) \sim N(\ln \mu(x_{bt}^*), 1)
\]  
(5.32)

**Condition 2:** If at least one brand has optimal zero product quantity, generate the utilities of brands with zero product quantity from the normal distribution

\[
\ln u(x_{bt}^* = 0) \sim N(\ln \mu(x_{bt}^*), 1), \forall b \in i'
\]  
(5.33)

Then generate the utilities of brands with positive quantity form the truncated normal distribution

\[
\ln u(x_{bt}^* \neq 0) \sim N(\ln \mu(x_{bt}^*), 1) \cdot I \left( \max_{i'} \{ \ln u(x_{bt}^* = 0) \} > \ln u(x_{bt}^* \neq 0) \right), \forall b \in i''
\]  
(5.34)

Because of the assumption that all items with the same brand name have the same errors \( \varepsilon_{bt} \) (see Equation (5.18)), the log latent utilities of those items not in the optimal set can be constructed by adding shared errors to the deterministic part of log utilities. Specifically, after generating the log latent utilities for all items in the optimal set, the error realizations for each brand can be obtained by

\[
\varepsilon_{bt} = \ln u(x_{bt}^*) - \left[ \ln \mu(x_{bt}^*) \right]
\]

\[
= \ln u(x_{bt}^*) - \left[ z_{bt}^* \beta + \alpha_x \ln (x_{bt}^* + 1) + \alpha_y \ln (T - p(x_{bt}^*)) \right]
\]  
(5.35)

Then the log utilities of items not included in the optimal set can be computed by adding the error term \( \varepsilon_{bt} \) obtained in Equation (5.35) to the determinist part of log latent utility.
\[
\ln u(x_{ibt}) = \ln \bar{u}(x_{ibt}) + \varepsilon_{ibt} \\
= z_{ibt} \beta + \alpha_x \ln (x_{ibt} + 1) + \alpha_T \ln \left( T - p(x_{ibt}) \right) + \varepsilon_{ibt}, \forall x_{ibt} \neq x^*_{ibt}
\] (5.36)

The detail estimation procedure of the proposed models is illustrated in Appendix C.

5.3.3. Data Augmentation for Error Terms

In this paper, the new variant of error augmentation illustrated in Chapter 2 is used to estimate the proposed models. Different from applications in literature, the error terms cannot be generated directly from a distribution because no-choice decisions are included in the proposed models. No-choice decisions reveal a partial ranking of all available items. There is no direct correspondence between observed data and the errors in the proposed models. Therefore, error terms are computed and retained after generating latent utilities associated with the observed choice. The rest of parameter estimation is implemented by reconstructing the realization of latent utilities given the retained errors and the candidate draws of parameters (e.g., \( \beta^{(j)}_h \), \( \alpha^{(j)}_{xh} \), \( \alpha^{(j)}_{Th} \), \( T^{(j)}_h \) in Equation (5.18) ) generated from the Metropolis-Hasting step. Different from Allenby et. al. (2003) in which the parameter estimation can be proceed conditional on the optimal quantities of brands, the candidate draws of parameters are accepted only if the rank of reconstructed latent utilities is consistent with the preference order implied by data.
5.4. Empirical Application

5.4.1 Data of consumer packaged goods

The proposed model is illustrated with a panel data of a product category of consumer packaged goods collected in an information acceleration study. Information acceleration (IA) is a measurement methodology suggested by Urban, et. al. (1996). The basic idea behind IA is to study consumer response (e.g., choice or perception) by “placing consumers in a multimedia virtual shopping environment that simulates the information that is available to consumers at the time they make a purchase decision” (Urban, et. al. 1996). Experimental design of IA is not limited in a single format. Researchers can design an experiment to meet particular research objectives.

This company would like to compete with its major competitor in a product category by introducing a new item that shares the same trademark with an established brand and provides similar product attributes as its major competitor. This marketing activity is known as line extension. The company conducts this IA study to predict the performance of new item and to evaluate the effectiveness of line extension. In order to maintain the confidentiality of this company, Brand A+ represents the new item introduced by this company, Brand A represents the items sharing the trademark with Brand A+, and brand B is the major competitor of brand A and has the similar product attributes as Brand A+. Brand C and D are other offerings in the market. Brand E represents the store brand.

The IA study conducted by this company consists of three steps: pretest, information acceleration, and posttest. First, 433 respondents are brought into a central site of the study. Respondents are asked to make choice among SKUs of established
brands (Brand A, B, C, D, E) in four pretest virtual shopping trips. Then, different sources (e.g., TV, magazine) of product information of new product (Brand A+) and established brands are simulated on computers in the step of information acceleration. Subject to an overall time limit (15 minutes in this study), respondents can access information in any order, revisit information source, and spend as much time as they want on any information source. The content of information (e.g., TV commercial of brand B) is same for all respondents, but not all information sources are available to respondents. After the step of information acceleration, respondents are asked to make choice again among SKUs of new product and established brands in six posttest shopping trips. Marketing merchandising variables shown in Table 5.1 are varied among shopping trips. The choice decision of each respondent occurs at the level of stock keeping unit (SKU). It implies that both brand and quantity decisions are involved in a choice task. Respondents are not forced to make choice because not all SKUs in the market are available in this study. In this dataset, 74 out of 433 respondents decide to not choose any item from the supermarket shelf at least once either in pre-IA or in post-IA shopping trips.

There are 45 SKUs in pre-IA shopping trips, and 51 SKUs in post-IA shopping trips. Product attributes of all SKUs are invariant in this study. Four major product attributes are used to differentiate offerings in this product category. The levels of product attributes are provided in Table 5.1. Attribute 1 and 2 are not provided by Brand A and are common to some SKUs of Brand B and Brand A+. Binary coding is applied to both product attributes and merchandising variables except price and product quantity.

Descriptive statistics of the data is provided in Table 5.2. Table 5.2(a) shows that Brand A, followed by Brand B, is the leading brand in this product category in the pre-IA
study. Table 5.2(b) shows that the new product Brand A+, the line extension of Brand A, takes the market share from Brand A even though the purchase frequencies of Brand A+ and Brand A together are account for 42.3% of total purchase in the post-IA study. Table 5.2(b) shows that the unit price of Brand A+ is higher than the unit price of Brand B even though Brand A+ and Brand B have similar product attributes.

The brand switching matrices are provided in Table 5.3. Table 5.3(a) shows that, in the pre-IA study, the consumers of Brand A tend to purchase Brand B or Brand E. Brand D and Brand B have strong competition. However, the brand switching matrix shown in Table 5.3(b) shows that Brand A+, Brand A, and Brand B compete against each other after Brand A+ is introduced to the market in the posttest study. In addition, the brand switching matrix shown in Table 5.3(c) shows that, right after information acceleration, 24.39% of respondents who choose Brand B and 16.34% of respondents who choose Brand A at the last pre-IA shopping trip switch to Brand A+ at the first post-IA shopping trip. Table 5.2(b) and Table 5.2(c) show that Brand A+ successfully attracts the consumers of Brand B right after information acceleration. However, Brand A+ does not keep these respondents in the rest of posttest shopping trips.

Brand-specific information from seven different media is available to respondents in the step of information acceleration. Media provided in this study includes television, magazine, coupon, the direct-to-consumer brochure (DTC brochure), web pages, and word of month, and sticker talker. The frequency of media exposures is reported in Table 5.4. No information sources of Brand C and Brand E are provided in this study since, in the reality, Brand C and E do not have many marketing activities in media. Printed advertisements of brands not in this product category are used as the filter in the
magazine. Coupon is included in the magazine, but cannot be used in shopping trips. Direct-to-consumer (DTC) brochure, word of mouth (WOM), and sticker talker only provide the product information of Brand A+. Word of mouth is given by telephone and only provides the recommendation of Brand A+.

Sticker talker is the sticker posted on the packaged of Brand A to introduce Brand A+ to the market. Sticker talker is only given in the last pre-IA shopping trip. In this study, sticker talker is considered as a medium of promoting Brand A+. Respondents are said to be exposed to sticker talker if respondents purchase Brand A with the sticker on the packaged in the last pre-IA shopping trip. Table 5.5 shows that sticker talker are given in the last pre-IA shopping trip of 211 respondents. Among these 211 respondents, only 80 respondents choose Brand A with sticker talker, and 12 respondents choose Brand A in the last pretest shopping trip and Brand A+ in the first posttest shopping trip.

Since media given in IA is varied among respondents and respondents decide which media they would like to review, some respondents do not have any media exposure. Table 5.6 reports the number of respondents who do not select media of each brand. Among 433 respondents, 19 respondents are never exposed to product information of all brands in the study.

5.4.2 Proposed models for the empirical study

The model developed in Section 5.3 is modified for analyzing the data illustrated in Section 5.4.1. Let $h$ denote the respondent and $n$ denote the number of treatments used in a study. Established Brand A to Brand E are available in a pretest study, and the new brand (Brand A+) is introduced in the posttest study. Assume that $m_{bh}$ is a vector contains
the binary outcome for brand-specific treatments, such as the TV commercial of brand \( b \), received by respondent \( h \). Since a consumer’s preferences and sensitivities to merchandising variables can be affected by media of any brand, two specifications of aggregating media exposures across brands are tested in this study.

The first specification is to standardize media exposures across all brands by assuming that each element in the vector \( m_{bh} \) contains the proportion of a medium’s exposure frequency summed across brands to the total media exposure. For example, the first element of \( m_{bh} \) for respondent \( h \) equals to \( \frac{\sum_b m_{1bh}}{\sum_b \sum_n m_{nbh}} \) for the first medium. The reason to standardize media exposures across all brands is that, in the empirical study, the frequencies of media exposures are concentrated on TV commercials, and media exposures are not balance among brands because some brands have more media choices (see Table 5.4). Standardizing media exposures may avoid concluding the media with high exposure frequency are more effective. Since standardizing media exposure can distort the data, the second specification assumes that each element in the vector \( m_{bh} \) contains the summation of media exposures across brands. For example, the first element of the vector \( m_{bh} \) equals to \( \sum_b m_{1bh} \) for the first medium.

Because Brand A+ is the line extension of Brand A and Brand A shares the trademark with Brand B, the intercept of Brand A+ is assumed to come from three sources: the product information of Brand A delivered by media, line extension, and other unobserved factors. That is,

\[
\beta_{01,h} = \beta_{1,h} + \gamma_h m_{1h} + \psi_h \beta_{02,h} \tag{5.37}
\]

where \( m_{1h} \) is a vector for the media of Brand A+ reviewed by respondent \( h \), \( \beta_{01,h} \) is the respondent \( h \)'s Brand A intercept evaluated in the pretest study, \( \gamma_h \) is the vector of media
effect, $\psi_h$ is the coefficient for the effect of line extension, and $\beta_{1,h}$ is the brand preference of brand 1 from other unobserved factors.

Given the multiplicative relationship between pretest and posttest measurements (Equation (5.26) and (5.27)), the posttest measurements are specified as

$$
\beta_{2h}^{(2)} = C_h \beta_{2h}^{(1)}
$$

$$
C_h = \text{diag} \left( 1 + \gamma_h m_{2h}, \cdots, 1 + \gamma_h m_{nh}, 1 + \theta_{1,h} m_{ch}, \cdots, 1 + \theta_{\text{var},h} m_{ch}, 1 + \theta_x m_{ch} \right)
$$

$$
\beta_{2h}^{(1)} = \left( \beta_{02,h}, \beta_{03,h}, \cdots, \beta_{06,h}, \beta_{z1,h}, \beta_{z2,h}, \cdots, \beta_{\text{var},h}, \beta_{ch} \right)^{1(\prime)}
$$

$$
\beta_{2h}^{(2)} = \left( \beta_{02,h}, \beta_{03,h}, \cdots, \beta_{06,h}, \beta_{z1,h}, \beta_{z2,h}, \cdots, \beta_{\text{var},h}, \beta_{ch} \right)^{2(\prime)}
$$

$$
\beta_{T,h}^{(2)} = \exp \left( \beta_{T,h}^{(2)\prime} \right) = \exp \left( \left( 1 + \theta_{T,h} M_h \right) \cdot \beta_{T,h}^{(1)\prime} \right)
$$

where $C_h$ is a diagonal matrix that contains media effects $\gamma_h$ of respondent $h$ for Brand A to Brand E and media effects $\theta_{ih}$ of respondent $h$ for all independent variables and ln$(x_{bh}^{(2)} + 1)$.

For the additive consumer heterogeneity (Equation (5.28) and (5.29)), the relationship between pretest measurements and posttest measurements is specified as:

$$
\beta_{2h}^{(2)} = \beta_{2h}^{(1)} + \Delta_h
$$

$$
\Delta_h = \left( \gamma_h m_{2h}, \cdots, \gamma_h m_{nh}, \theta_{1,h} m_{ch}, \cdots, \theta_{\text{var},h} m_{ch}, \theta_x m_{ch} \right)
$$

$$
\beta_{2h}^{(1)} = \left( \beta_{02,h}, \beta_{03,h}, \cdots, \beta_{06,h}, \beta_{z1,h}, \beta_{z2,h}, \cdots, \beta_{\text{var},h}, \beta_{ch} \right)^{1(\prime)}
$$

$$
\beta_{2h}^{(2)} = \left( \beta_{02,h}, \beta_{03,h}, \cdots, \beta_{06,h}, \beta_{z1,h}, \beta_{z2,h}, \cdots, \beta_{\text{var},h}, \beta_{ch} \right)^{2(\prime)}
$$

$$
\beta_{T,h}^{(2)} = \exp \left( \beta_{T,h}^{(2)\prime} \right) = \exp \left( \left( 1 + \theta_{T,h} M_h \right) \cdot \beta_{T,h}^{(1)\prime} \right)
$$

where $\Delta_h$ is a column vector that contains media effects $\gamma_h$ of respondent $h$ for Brand A to Brand E and the media effect $\theta_{ih}$ of respondent $h$ for all independent variables and ln$(x_{bh}^{(2)} + 1)$.
Same priors are applied to models with multiplicative and additive specifications.

The prior distributions are defined as follows:

\[
\begin{align*}
\beta_{1,h} & \sim N(0,10); \psi_h \sim N(0,10) \\
\beta_h^{(1)*} & \sim N(\bar{\beta}, V_\beta); \tilde{\beta} \sim N(\mu_0 = 0, V_0 = 100); V_\beta \sim IW(\psi_0 = 27, G_0 = 27I) \\
T_h & \sim N(\bar{T} = 10, \sigma^2 = 50) \quad (5.40) \\
\gamma_h & \sim N(\bar{\gamma} = 0, V_\gamma = 50I) \\
\theta_{z,h} & \sim N(\bar{\theta} = 0, V_\theta = 50I)
\end{align*}
\]

5.4.3. Location identification of proposed models

It has been known that multinomial models have both location and scale identification problems. That is, the likelihood of a multinomial model is invariant to location or scale transformation. The conventional approach to identify the location of a multinomial probit model is to fix the coefficient of one brand intercept to zero, and the parameter estimates of intercept of brands are explained as the contrast to the base brand. The scale identification of a multinomial probit model can be achieved by either setting one diagonal element of covariance matrix to 1 or setting one element of parameter vector to some fixed number (see Rossi, et. al. 2005).

For the proposed model for the empirical application, the empirical identification of scale parameter is solved by fixing the covariance matrix of latent utilities as an identity matrix. In addition, the assumption of shared random-effect realizations, the parameterization of multiplicative and additive relationships between pretest and posttest measurements, and the parameterization for the line extension effect allow consumer preferences to be estimated without fixing the intercept of one brand to zero.
Specifically, let $x_{ih}^{*}$ denote the optimal quantity of brand $i$ for respondent $h$ at the shopping trip $t$. The $b^{th}$ row vector of design matrix $z_{ht}^{*}$ contain brand intercept, merchandising variables such as deal and display, $\ln(x_{ih}^{*}+1)$, and $\ln(T_h-p(x_{ih}^{*}))$ for optimal packsize $x^*$ of brand $b$ purchased by respondent $h$ at the shopping trip $t$. The pretest translated log Cobb-Douglas utility function for the optimal quantities of all brands in pre-IA study is

$$\ln u(x_{ht}^{(1)*}) = z_{ht}^{(1)*} \beta_{h}^{(1)} + e_{ht}^{(1)*}$$ (5.41)

The posttest translated log Cobb-Douglas utility function for the optimal quantities of all brands in post-IA is

$$\ln u(x_{ht}^{(2)*}) = z_{ht}^{(2)*} \beta_{h}^{(2)} + e_{ht}^{(2)*}$$ (5.42)

Let $z_{ht}^{(2)}$ denote the design matrix $z_{ht}^{(2)*}$ excluding intercepts. According to the multiplicative specification defined in Equation (5.38) and the brand preference to brand A+ defined in Equation (5.37), the posttest translated log Cobb-Douglas utility function can be rewritten as

$$\ln u(x_{ht}^{(2)*}) = z_{ht}^{(2)*} \beta_{h}^{(2)} + e_{ht}^{(2)*}$$

(5.43)

According to the additive specification defined in Equation (5.39) and line extension effect defined in Equation (5.37), the posttest translated log Cobb-Douglas utility function can be rewritten as
\[
\ln u(x_{it}^{(2)*}) = z_{it}^{(2)*} \beta_{h}^{(2)} + e_{it}^{(2)*}
\]

\[
= \begin{bmatrix}
\beta_{1,h} + \gamma_{1} m_{1,h} + \psi_{1} \beta_{01,h}^{(1)} \\
\beta_{02,h}^{(1)} + \gamma_{2} m_{2,h} \\
\vdots \\
\beta_{06,h}^{(1)} + \gamma_{6} m_{6,h}
\end{bmatrix}
+ \begin{bmatrix}
\beta_{z_{1},h} + \theta_{z_{1}} M_{h} \\
\beta_{z_{2},h} + \theta_{z_{2}} M_{h} \\
\vdots \\
\exp\left( \beta_{T,h}^{(1)*} + \theta_{T} M_{h} \right)
\end{bmatrix}
+ e_{it}^{(2)*}
\] (5.44)

In both multiplicative and additive specifications, brand intercepts of pretest translated log Cobb-Douglas utility function are constrained by both pretest and posttest observations because both pretest and posttest observations provide information to estimate pretest measurements \( \beta_{1,h}^{(1)} \). The posttest likelihood will be altered if adding an arbitrary constant to the posttest translated log Cobb-Douglas utility of each brand. Therefore, the location identification problem is not an issue in the proposed model given the assumption of shared random-effect realizations, the specifications of multiplicative and additive relationships between pretest and posttest measurements, and the parameterization for line extension effect.

5.5. Results

5.5.1. Model Comparison

In addition to the model with multiplicative heterogeneity and the model with additive consumer heterogeneity, two alternative models are proposed for the model comparison. The summary of models and their log marginal densities are reported in Table 5.7.

The first alternative model (baseline model) is to analyze pre-IA and post-IA data with a multinomial probit model separately. This model is considered to be the standard
method of studying media effects in a pre-post study. Media effects are studied by comparing parameter estimates of pre-IA data and post-IA data. That is, the pre-IA data and the post-IA data are analyzed respectively with the following model:

\[ Pr(x_{\text{post}}^* = x_{\text{pre}}^*) > \max \{ \ln u(x_{\text{pre}}^*) \forall b' \neq b \} \]

\[ Pr(0) = Pr(\max_i \{ \ln u(x_{\text{pre}}^* = 0) \}) > \max_i \{ \ln u(x_{\text{pre}}^* \neq 0) \} \] \{i' \cup i^* \} = \{i\} \quad (5.45)

\[ u(x_{\text{obs}}^*) = z_{\text{obs}} \beta_h + \alpha_x \ln(x_{\text{obs}}^* + 1) + \alpha_T \ln(T - p(x_{\text{obs}}^*)) + \epsilon_{\text{obs}} + \epsilon_{\text{obs}} \sim N(0,1) \]

\[ (\beta_h, \alpha_x, \alpha_T)^T \sim N(\mu, V) \quad (5.46) \]

The other alternative model is specified with two distinct random-effect distributions for pretest and posttest data. Media effects are studied by assuming that the posttest consumer heterogeneity can be explained by media exposures. The likelihood specified in Equation (5.17) and (5.18) is used to analyze the data. The distribution of consumer heterogeneity for the pretest measurements is specified as follows:

\[ \beta_{(1)^*} = [\beta_{(1),h}^{(1)} \beta_{(1),b}^{(1)} \cdots \beta_{(1),i}^{(1)} \beta_{(1),j}^{(1)} \cdots \alpha_{x,h}^{(1)} \alpha_{T,h}^{(1)^*}]^T \sim N(\overline{\beta}^{(1)}, V_{\beta}^{(1)}) \quad (5.47) \]

\[ \overline{\beta}^{(1)} = [\beta_{(1),h}^{(1)} \beta_{(1),b}^{(1)} \cdots \beta_{(1),i}^{(1)} \beta_{(1),j}^{(1)} \cdots \alpha_{x}^{(1)} \alpha_{T}^{(1)^*}]^T \]

The distribution of consumer heterogeneity for the posttest measurements is specified as follows:

\[ \beta_{(2)^*} = [\beta_{(2),h}^{(2)} \beta_{(2),b}^{(2)} \beta_{(2),i}^{(2)} \cdots \beta_{(2),j}^{(2)} \beta_{(2),i}^{(2)} \beta_{(2),j}^{(2)} \cdots \alpha_{x,h}^{(2)} \alpha_{T,h}^{(2)^*}]^T \sim N(\overline{\beta}^{(2)}, V_{\beta}^{(2)}) \quad (5.48) \]

where
Both pre-post intercepts of brand B in these two alternative models are fixed to zero for the location identification of the model.

All models except baseline model and multiplicative model, models with standardized media ($m_{kh}=(\Sigma b_{mk}k_b)/(\Sigma b_{mn}n_b)$ for the $k$th media) are used to analyze the data. Since the multiplicative model has the best fit, the multiplicative model with non-standardized media exposure ($m_{kh}=(\Sigma b_{mk}k_b)$ for the $k$th media) is used to compare the result of different specifications of media exposure.

5.5.2. Coefficient Estimates

As shown in Table 5.7, the model with multiplicative specification and non-standardized media exposure (Model 4), except the baseline model, fits the data. The baseline model has the best fit among all five models because the baseline model is relatively flexible. However, the baseline model cannot be used to study the media effects. Since the multiplicative model with non-standardized media exposure has the best fit, the following discussion focuses on the results of the multiplicative model.
The posterior estimates of $\bar{\beta}$ and $V_\beta$ for four models are reported in Table 5.8 and Table 5.9. In Table 5.8, parameter estimates of $\bar{\beta}$ of baseline pre-IA model, two random-effect model, multiplicative, and additive models provides consistent results to consumer preferences in pre-IA study. Parameter estimates of baseline model in Table 5.8 also show that product attribute 1 is disliked by respondents in pre-IA, but is preferred by respondents in post-IA. When the levels of independent variables of Brand A and Brand B are the same, Brand A is prefer to Brand B in pre-IA, and Brand B is prefer to Brand A in post-IA. The basic level of product attribute 4 is most preferred. Product attribute 1, 2, and 3 are not preferred by respondents. Merchandising variables (deal, display, shelf talker, and bonus pack) either have negative impacts or have few impacts in pre-IA study. Respondents do gain utilities by purchasing large quantities of products.

The difference between pre and post brand intercepts can be observed from Figures 5.1 to 5.4. Figure 5.1 presents box plots of the pre-post posterior mean of brand intercepts. Figure 5.1 shows that, when the levels of independent variables of all brands are the same, some consumers have extreme brand preferences after information acceleration. Figure 5.2 provides scatter plots of pre-post intercepts of Brand A, B, and D. As shown in Figure 5.2, pre-post brand intercepts are positive correlated. Figure 5.3 provides scatter plots of post brand intercept of Brand A+ and pre brand intercepts of all established brands. Figure 5.4 provides scatter plots of post brand intercepts of Brand A+ and all established brands. Both figures show that the number of respondents who have positive intercept of Brand A and positive intercept of Brand A+ is more than the number of respondents who have positive intercept of Brand B and positive intercept of Brand A+. 
Line extension effect can be studied from Equation (5.37) and Figure 5.5. In Equation (5.37), $\beta_{l,h}$ represents the effect to the intercept of Brand A+ that is not explained by information sources of Brand A+ and line extension. $\psi_h \beta^{(1)}_{02,h}$ is for line extension effect. Scatter plots in Figure 5.5 are used to diagnose three components of Brand A+ intercept. Figure 5.5 shows that the intercept of Brand A+ is positive correlated with the effect of line extension ($\psi_h \beta^{(1)}_{02,h}$). Information sources of Brand A+ have a small impact on the intercept of Brand A+.

Figure 5.6 provides histograms for the difference of posterior mean of pre-post consumer sensitivities to merchandising variables: deal, display, shelf talker, and bonus pack. Since histograms of merchandising variables spread out widely, Figure 5.6 also reveals that consumers respond to merchandising variables are very heterogeneous after information acceleration.

Product attribute 1 is offered by Brand A+, Brand B, and Brand D. Product attributes 2 are offered by Brand A+ and Brand B. The manufacture of Brand A introduces Brand A+ as a line extension to compete with Brand B. Figure 5.7 and Figure 5.8 provide the histogram of the difference of pre-post posterior mean of consumer preferences to product attributes. Figure 5.7 shows more respondents have preferences to product attribute 1 after information acceleration. More respondents are averse to product attribute 3 after information acceleration. Figure 5.8 shows that most respondents prefer to the basic level of product attribute 4 and become averse to other attribute levels after information acceleration. The difference of pre-post consumer preferences to quantity and price sensitivities are shown in Figure 59. Respondents prefer to large quantity of
product after information acceleration. Most respondents become less price sensitive after information acceleration.

The coefficient estimates of media effects on brands preference ($\gamma$) for the multiplicative specification and the additive specification are presented in Table 5.10. Posterior estimates of $\nu_{r}$ presented in Table 5.11. Media effects on the intercept of each brand are presented in Figure 5.10, 5.11, 5.12, and 5.13.

Figure 5.10 provide a box plot for the posterior mean of $\gamma_{h}$ for all information sources of Brand A+. It shows that commercial, magazine, the first page of DTC brochure, web page containing information of product attribute 1 and web page 6 have positive impacts on the intercept of Brand A+. The first page of DTC brochure with product information (Brochure info-page 1), the main web page of Brand A+, and sticker talker have negative impact on the intercept of Brand A+.

Posterior estimates of $\gamma_{h,m_{0b,h}}$ is used to diagnose the aggregate media effects on the intercept of Brand b when the multiplicative specification is given. Since the sign of $\beta_{0b,h}^{(1)}$ can reverse the impact of media, the box plot of $\gamma_{h}\beta_{0b,h}^{(1)}$ is also provided. Figures 5.11 to 5.13 reports box plots of posterior mean of $\gamma_{h}$ and $\gamma_{h}\beta_{0b,h}^{(1)}$, histogram of the difference of pre-post brand intercept ($\beta_{0b,h}^{(2)} - \beta_{0b,h}^{(1)}$), and histogram of the aggregate media effect ($\gamma_{d,m_{0b,h}}$). Since no information sources of Brand C and E are provided in this study, only media effects of Brand A, B, and D are discussed.

Histogram of the difference of pre-post intercepts of Brand A ($\beta_{02,h}^{(2)} - \beta_{02,h}^{(1)}$) is given in Figure 5.11(a). Histogram of the aggregate media effect of Brand A is given in Figure 5.11(b). Both plots show that media have positive impacts on the post intercept of Brand
A. Figure 5.11(c) provides the box plot of media effect $\gamma_h$ for all information sources of Brand A. It shows that commercials, magazine, and coupon in magazine have positive impacts on intercept of Brand A. Figure 5.11(d) shows that pre brand intercept $\beta_{02,h}^{(i)}$ makes some respondents’ media effects $\gamma_h$ for all information sources either become extreme or become neural.

Histogram of the difference of pre-post intercepts of Brand B ($\beta_{03,h}^{(2)} - \beta_{03,h}^{(1)}$) is given in Figure 5.12(a). Histogram of the aggregate media effect of Brand B is given in Figure 5.12(b). Both plots show that media have positive impacts on post intercept of Brand B. Figure 5.12(c) provides the box plot of media effect $\gamma_h$ for all information sources of Brand B. It shows that commercials, magazine, and coupon in magazine have positive impacts on intercept of Brand B. The main web page of Brand B has a negative impact on Brand B. Figure 5.12(d) shows that pre brand intercept $\beta_{03,h}^{(i)}$ make some respondents’ media effects $\gamma_h$ for all information sources either become extreme or become neural.

Histogram of the difference of pre-post intercepts of Brand D ($\beta_{05,h}^{(2)} - \beta_{05,h}^{(1)}$) is given in Figure 5.13(a). Histogram of the aggregate media effect of Brand D is given in Figure 5.13(b). Figure 5.13(a) shows that media have negative impact on the intercept of Brand D even thought positive aggregate media impacts are observed in Figure 5.13(b). It is because most respondents’ pre intercepts of Brand D ($\beta_{05,h}^{(i)}$) are negative.

Figure 5.13(c) provides the box plot of media effect $\gamma_h$ for all information sources of Brand D. It shows that media sources, except the web pages containing product information of product attribute 4 with level 1 and level 2, have positive effects. Figure
5.13(d) shows that pre brand intercept $\beta_{0.5,h}^{(1)}$ makes some respondents’ media effects $\gamma_h$ for all information sources either become extreme or become neural.

The coefficient estimates of the media interactions with merchandising variables ($\bar{\theta}$ ) for the multiplicative specification and the additive specification are provided in Table 5.12. As seen in Table 5.12, except commercial-partial and the third page of DTC brochure, the effects of media interactions ($\bar{\theta}$ ) are not very significant in the multiplicative model. Table 5.13 shows that, the posterior estimates of $V_{\theta}$ of multiplicative models have great amount of consumer heterogeneity in media effects on merchandising variables.

$\theta_{z,h}m_{z,h}$ represents aggregate media effects in the multiplicative specification. Thus, in multiplicative specification, small effect of one medium can have great aggregate media effects. $\theta_{T,h}\beta_{T,h}^{(1)r}$ represents media interaction between media and merchandising variables.

Figure 5.14 provides the box plots of aggregate media effect for each merchandising variables. Greater heterogeneity is observed in the posterior distribution of aggregate media effects ($\theta_{z,h}m_{z,h}$). Only the medium of the aggregate media effect on quantity ($\log(x+1)$) is positive. Because the posterior distributions of aggregate media effect $\theta_{z,h}m_{z,h}$ and $\theta_{z,h}$ are almost identical for all merchandising variables except product quantity (see Figure 5.15(a) and 5.15(b) as examples), only the media effect on product attribute, $\ln(x+1)$, and price sensitivity ($\ln(T-p(x))$) are reported.

Figure 5.15(a) provides box plots for media interactions between product attribute 1 and all information sources. Figure 5.15(b) provides box plots for the posterior mean of
\( \theta_{z,h} \) for product attribute 1. Figure 5.15(c) provides histogram of aggregate media effects of product attribute 1. Figure 5.15 shows that, even though the posterior distributions of \( \theta_{z,h} \rho^{(1)}_{z,h} \) and \( \theta_{z,h} \) are centered at zero for all media and the distribution of aggregate media effects \( (\theta_{z,h} \beta_{z,h} m_{z,h}) \) is symmetric, most respondents’ post preferences to product attribute 1 are increased after information acceleration (see Figure 5.7(a)).

Figure 5.16(a) provides box plots for media interactions between \( \ln(x+1) \) and all information sources. Figure 5.16(b) provides box plots for the posterior mean of \( \theta_{z,h} \) for \( \ln(x+1) \). Figure 5.16(c) provides histogram of aggregate media effects of \( \ln(x+1) \). Figure 5.16(b) shows that commercials, magazine, and coupon in magazine have positive media effects. Figure 5.16(c) shows that most of respondents have positive aggregate media effects on \( \ln(x+1) \). The positive impact of information acceleration on \( \ln(x+1) \) can also be supported by Figure 5.9(a).

Note that the media interaction between merchandising variables and \( \ln(T-p(x)) \) cannot be studied in this specification because of the exponential transformation

\[
\beta_{T,h}^{(2)} = \exp(\beta_{T,h}^{(2)*}) = \exp\left((1 + \theta_{T,h} m_{z,h}) \cdot \beta_{T,h}^{(1)*}\right).
\]

Figure 5.17(a) provides box plots for \( \theta_{T,h} \beta_{T,h}^{(1)*} \). Figure 5.17(b) provides box plots for the posterior mean of \( \theta_{T,h} \) for \( \ln(T-p(x)) \). Figure 5.17(c) provides histogram of aggregate media effects of \( \ln(T-p(x)) \). It shows that respondents are less price sensitive after information acceleration even though the posterior distributions of \( \theta_{T,h} \beta_{T,h}^{(1)*} \) and \( \theta_{T,h} \) are centered at zero, and extreme impacts of \( \theta_{T,h} \beta_{T,h}^{(1)*} \) are observed.

In general, the multiplicative specification suggests that a small effect of each medium can lead to a large aggregate effect across media. This argument is supported by
Figure 5.15(c), Figure 5.16(c), and Figure 5.17(c). It also shows that most respondents’ sensitivities and preferences to merchandising variables are difficult to be altered by media in short time period.

5.6. Conclusion

A demand model for consumer packaged goods is developed in this paper to deal with three issues in choice modeling. The first issue relates to the effect of multiple treatments and consumer preference changes for data collected in a pre-post study. Treatment effects and consumer preference change are studied by relating pretest and posttest measurements (e.g., brand preference, price sensitivity) to same random-effect realizations, and by specifying an explicit relationship between pretest and posttest measurements. The method proposed in this paper can be used to study multiple treatment effects in a pre-post study, and can be used to study consumer brand preference change.

The second issue relates to line extension that is widely adopted in marketing practice. However, few methods in marketing are developed to evaluate the effect of line extension. In this paper, a consumer’s preference to a new brand is assumed to come from media containing product information of new brand, line extension effect, and other unobserved factors. By decomposing a consumer’s preference to a new brand, researchers can diagnose the effect of line extension, and study the factors that have great impact on consumer preferences to new brand.

The third issue relates to no-choice decisions occurred at the level of stock keeping unit. The challenge of modeling no-choice decisions is that no-choice decisions reveal a
partial ranking among all available brand-pack combination. Translated Cobb-Douglas utility is used to accommodate no-choice decisions. A direct evaluation procedure and a new variant of data augmentation are used to overcome estimation difficulties associated with no-choice decisions.

Data collected in an information acceleration study are analyzed with the proposed model. The update and carry-over of consumer preferences are studied by specifying multiplicative and additive relationships between pre-post measurements. The results of empirical study show that posttest consumer preferences are updated according to multiplicative relationship. Media effects are a function of pretest consumer preferences.

Results of empirical study show that consumers have very different reactions to media. Information acceleration can make posttest consumer preferences go to extreme. In general, commercials and magazine of new product, Brand A+, have positive effects on consumer preferences to Brand A+. However, consumer preferences to new brand are attributed to line extension effect, not media effect. Besides, consumers respond to media in a very different way. Preferences and sensitivities to brands and merchandising variables of around 50% respondents are not affected by media. It shows that consumer preferences cannot be altered by media in a short run.

This paper provides a method to study multiple treatments for data collected in a pre-post study and to study no-choice decisions often observed in practice. Because of the limitation of data, only short-term effects can be identified in this research. Future research can be extended to the applications of preference changes in multiple stages, and long-term treatment effects.
Figure 5.1 Box plots for the posterior mean of brand intercept of multiplicative model: (a) pre-IA $\beta$; (b) post-IA $\beta$
Figure 5.2 Scatter plots for pre-post posterior mean of brand intercept of multiplicative model: (a) Brand A; (b) Brand B; (c) Brand D
Figure 5.3 Scatter plots for posttest brand intercept of Brand A+ and pretest brand preference intercept of established brands
Figure 5.4 Scatter plots for posttest brand intercept of Brand A+ and posttest brand intercept of established brands
Figure 5.5 Components of consumer preferences intercept of Brand A+ ($\beta_{01,h}$): (a) $\beta_{01,h}$ v.s. $\beta_{1,h}$; (b) $\beta_{01,h}$ v.s. $\gamma_{h_1,h}$; (c) $\beta_{01,h}$ v.s. $\psi_{h_2,h}\beta_{02,h}$
Figure 5.6 Histogram for the difference of pre-post posterior mean of consumer sensitivities to marketing merchandising variables of multiplicative model: (a) deal; (b) display; (c) shelf talker; (d) bonus pack
Figure 5.7 Histogram for the difference of pre-post posterior mean of consumer preferences to product attributes 1, 2 and 3 of multiplicative model: (a) product attribute 1; (b) product attribute 2; (c) product attribute 3
Figure 5.8 Histogram for the difference of pre-post posterior mean of consumer preferences to product attributes 4 of multiplicative model: (a) level 1; (b) level 2; (c) level 3
Figure 5.9 Histogram for the difference of pre-post posterior mean of consumer preference to quantity ($\beta_{x,h}$) and the outside goods ($\beta^*_{T,h}$) of multiplicative model: (a) product quantity $\ln(x+1)$; (b) the outside goods $\ln(T-p(x))$ (after taking log transformation)
Figure 5.10 Box plots for the posterior mean of $\gamma_h$ for all information sources of Brand A+
Figure 5.11 Media effects on the intercept of Brand A: (a) $\beta_{02,h}^{(2)} - \beta_{02,h}^{(1)}$; (b) $\gamma_h m_{02,h}$; (c) $\gamma_h$; (d) $\gamma_h \beta_{02,h}^{(1)}$;
Figure 5.11 Continued

(c)

(d)
Figure 5.12 Media effects on the intercept of Brand B: (a) $\beta - \beta_1$; (b) $\gamma \cdot m$; (c) $\gamma_2$; (d) $\gamma_3 \cdot \rho_0$.

(Continued)
Figure 5.12 Continued

(c)

(d)
Figure 5.13 Media effects on the intercept of Brand D: (a) $\beta_{05,h}^{(2)} - \beta_{05,h}^{(1)}$; (b) $\gamma_h m_{05,h}$; (c) $\gamma_h$; (d) $\gamma_h \beta_{05,h}^{(1)}$
Figure 5.13 Continued

(c)

(d)
Figure 5.14 Aggregate media effects ($\theta_{z,k} M_k$)
Figure 5.15 Media effects on consumer preferences to product attribute 1:
(a) $\theta_{z,h}^{(i)}$; (b) $\theta_z$; (c) $\theta_{z,h} m_h$
Figure 5.15 Continued
Figure 5.16 Media effects on consumer preferences to $\ln(x+1)$: (a) $\theta_{z,h} \beta^{(l)}_{z,h}$; (b) $\theta_{z,h}$; (c) $\theta_{z,h} m_{z,h}$.
Figure 5.16 Continued

(c)
Figure 5.17 Media effects on consumer sensitivities to $\ln(T-p(x))$: (a) $\theta_{T,h} \beta^{(1)}_{T,h}$; (b) $\theta_{T,h}$; (c) $\theta_{z,h} m_{z,h}$
Figure 5.17 Continued

![Histogram of theta*media](image)

(c)
# Table 5.1 Levels of independent variables

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Marketing merchandising variables</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Product attributes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Product quantity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Price</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Brand A</td>
<td>Brand B</td>
<td>Brand C</td>
<td>Brand D</td>
<td>Brand E</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>shopping trip 1</td>
<td>193</td>
<td>101</td>
<td>15</td>
<td>54</td>
<td>26</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>shopping trip 2</td>
<td>182</td>
<td>78</td>
<td>19</td>
<td>59</td>
<td>39</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>shopping trip 3</td>
<td>137</td>
<td>102</td>
<td>17</td>
<td>65</td>
<td>56</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>shopping trip 4</td>
<td>165</td>
<td>85</td>
<td>17</td>
<td>47</td>
<td>66</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>677</td>
<td>366</td>
<td>68</td>
<td>225</td>
<td>187</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percentage</td>
<td>44.45%</td>
<td>24.03%</td>
<td>4.46%</td>
<td>14.77%</td>
<td>12.28%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Price per unit</td>
<td>0.16</td>
<td>0.21</td>
<td>0.16</td>
<td>0.16</td>
<td>0.12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(a)

<table>
<thead>
<tr>
<th></th>
<th>Brand A+</th>
<th>Brand A</th>
<th>Brand B</th>
<th>Brand C</th>
<th>Brand D</th>
<th>Brand E</th>
</tr>
</thead>
<tbody>
<tr>
<td>shopping trip 1</td>
<td>59</td>
<td>73</td>
<td>135</td>
<td>15</td>
<td>31</td>
<td>64</td>
</tr>
<tr>
<td>shopping trip 2</td>
<td>55</td>
<td>122</td>
<td>85</td>
<td>14</td>
<td>72</td>
<td>38</td>
</tr>
<tr>
<td>shopping trip 3</td>
<td>46</td>
<td>127</td>
<td>106</td>
<td>15</td>
<td>46</td>
<td>48</td>
</tr>
<tr>
<td>shopping trip 4</td>
<td>91</td>
<td>61</td>
<td>114</td>
<td>12</td>
<td>50</td>
<td>59</td>
</tr>
<tr>
<td>shopping trip 5</td>
<td>71</td>
<td>71</td>
<td>119</td>
<td>12</td>
<td>52</td>
<td>55</td>
</tr>
<tr>
<td>shopping trip 6</td>
<td>87</td>
<td>110</td>
<td>63</td>
<td>17</td>
<td>47</td>
<td>58</td>
</tr>
<tr>
<td>Total</td>
<td>409</td>
<td>564</td>
<td>622</td>
<td>85</td>
<td>298</td>
<td>322</td>
</tr>
<tr>
<td>Percentage</td>
<td>17.78%</td>
<td>24.52%</td>
<td>27.04%</td>
<td>3.70%</td>
<td>12.96%</td>
<td>14.00%</td>
</tr>
<tr>
<td>Price per unit</td>
<td>0.27</td>
<td>0.18</td>
<td>0.20</td>
<td>0.17</td>
<td>0.16</td>
<td>0.12</td>
</tr>
</tbody>
</table>

(b)

Table 5.2 Descriptive statistics: (a) the frequency table for brand choice in pre-IA study; (b) the frequency table for brand choice in post-IA study
<table>
<thead>
<tr>
<th>Shopping trip t-1</th>
<th>Brand A</th>
<th>Brand B</th>
<th>Brand C</th>
<th>Brand D</th>
<th>Brand E</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brand A</td>
<td>333</td>
<td>78</td>
<td>52</td>
<td>70</td>
<td>96</td>
</tr>
<tr>
<td></td>
<td>52.94%</td>
<td>12.40%</td>
<td>8.27%</td>
<td>11.13%</td>
<td>15.26%</td>
</tr>
<tr>
<td>Brand B</td>
<td>72</td>
<td>306</td>
<td>117</td>
<td>132</td>
<td>134</td>
</tr>
<tr>
<td></td>
<td>9.46%</td>
<td><strong>40.21%</strong></td>
<td>15.37%</td>
<td>17.35%</td>
<td>17.61%</td>
</tr>
<tr>
<td>Brand C</td>
<td>31</td>
<td>117</td>
<td>127</td>
<td>113</td>
<td>107</td>
</tr>
<tr>
<td></td>
<td>6.26%</td>
<td>23.64%</td>
<td><strong>25.66%</strong></td>
<td>22.83%</td>
<td>21.62%</td>
</tr>
<tr>
<td>Brand D</td>
<td>59</td>
<td>128</td>
<td>109</td>
<td><strong>224</strong></td>
<td>117</td>
</tr>
<tr>
<td></td>
<td>9.26%</td>
<td>20.09%</td>
<td>17.11%</td>
<td><strong>35.16%</strong></td>
<td>18.37%</td>
</tr>
<tr>
<td>Brand E</td>
<td>49</td>
<td>128</td>
<td>107</td>
<td>118</td>
<td><strong>175</strong></td>
</tr>
<tr>
<td></td>
<td>8.49%</td>
<td>22.18%</td>
<td>18.54%</td>
<td>20.45%</td>
<td><strong>30.33%</strong></td>
</tr>
</tbody>
</table>

(a)

<table>
<thead>
<tr>
<th>Shopping trip t-1</th>
<th>Brand A+</th>
<th>Brand A</th>
<th>Brand B</th>
<th>Brand C</th>
<th>Brand D</th>
<th>Brand E</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brand A+</td>
<td>129</td>
<td>76</td>
<td>58</td>
<td>7</td>
<td>17</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>42.16%</td>
<td>24.84%</td>
<td>18.95%</td>
<td>2.29%</td>
<td>5.56%</td>
<td>6.21%</td>
</tr>
<tr>
<td>Brand A</td>
<td>62</td>
<td>238</td>
<td>67</td>
<td>9</td>
<td>23</td>
<td>34</td>
</tr>
<tr>
<td></td>
<td>14.32%</td>
<td><strong>54.97%</strong></td>
<td>15.47%</td>
<td>2.08%</td>
<td>5.31%</td>
<td>7.85%</td>
</tr>
<tr>
<td>Brand B</td>
<td>89</td>
<td>80</td>
<td><strong>278</strong></td>
<td>4</td>
<td>55</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>16.54%</td>
<td>14.87%</td>
<td><strong>51.67%</strong></td>
<td>0.74%</td>
<td>10.22%</td>
<td>5.95%</td>
</tr>
<tr>
<td>Brand C</td>
<td>11</td>
<td>7</td>
<td>7</td>
<td><strong>32</strong></td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>17.19%</td>
<td>10.94%</td>
<td>10.94%</td>
<td><strong>50.00%</strong></td>
<td>7.81%</td>
<td>3.13%</td>
</tr>
<tr>
<td>Brand D</td>
<td>18</td>
<td>27</td>
<td>33</td>
<td>3</td>
<td><strong>143</strong></td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>7.44%</td>
<td>11.16%</td>
<td>13.64%</td>
<td>1.24%</td>
<td><strong>59.09%</strong></td>
<td>7.44%</td>
</tr>
<tr>
<td>Brand E</td>
<td>26</td>
<td>37</td>
<td>23</td>
<td>8</td>
<td>16</td>
<td><strong>142</strong></td>
</tr>
<tr>
<td></td>
<td>10.32%</td>
<td>14.68%</td>
<td>9.13%</td>
<td>3.17%</td>
<td>6.35%</td>
<td><strong>56.35%</strong></td>
</tr>
</tbody>
</table>

(b)

(Continued)

Table 5.3 Brand switching matrices: (a) pre-IA study; (b) post-IA study; (c) the last shopping trip of pre-IA study and the first shopping trip of post-IA study.
Table 5.3 Continued

<table>
<thead>
<tr>
<th>The last shopping trip of pre-IA study</th>
<th>The first shopping trip in post-IA study</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Brand A+</td>
</tr>
<tr>
<td>Brand A</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Brand B</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Brand C</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Brand D</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Brand E</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(c)
<table>
<thead>
<tr>
<th>Source of Information</th>
<th>Brand A+</th>
<th>Brand A</th>
<th>Brand B</th>
<th>Brand D</th>
</tr>
</thead>
<tbody>
<tr>
<td>TV</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.Commercial-partial</td>
<td>145</td>
<td>286</td>
<td>310</td>
<td>302</td>
</tr>
<tr>
<td>2.Commercial-complete</td>
<td>272</td>
<td>272</td>
<td>257</td>
<td>261</td>
</tr>
<tr>
<td>Magazine</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.Magazine</td>
<td>124</td>
<td>248</td>
<td>255</td>
<td>253</td>
</tr>
<tr>
<td>4.Coupon-page</td>
<td>123</td>
<td>257</td>
<td>236</td>
<td>235</td>
</tr>
<tr>
<td>Direct-to-Consumer Brochure</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.Brochure-page1</td>
<td>58</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6.Brochure-page2</td>
<td>43</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>7.Brochure-page3</td>
<td>37</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>8.Brochure-coupon</td>
<td>42</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>9.Brochure info-page1</td>
<td>35</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10.Brochure info-page2</td>
<td>17</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Web pages</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.Web page 1 for Brand A+: Main</td>
<td>14</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>12.Web page 2 for Brand A+</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>13.Web page 3 for Brand A+</td>
<td>11</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>14.Web page 4 for Brand A+</td>
<td>11</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>15.Web page 5 for Brand A+: Attribute 1</td>
<td>12</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>16.Web page 6 for Brand A+</td>
<td>12</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>17.Web page 7 for Brand A+</td>
<td>12</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>18.Web page 1 for Brand A: Main</td>
<td>0</td>
<td>42</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>19.Web page 2 for Brand A: Products</td>
<td>0</td>
<td>33</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>20.Web page 1 for Brand B: Main</td>
<td>0</td>
<td>0</td>
<td>90</td>
<td>0</td>
</tr>
<tr>
<td>21.Web page 2 for Brand B</td>
<td>0</td>
<td>0</td>
<td>68</td>
<td>0</td>
</tr>
<tr>
<td>22.Web page 3 for Brand B: Product</td>
<td>0</td>
<td>0</td>
<td>57</td>
<td>0</td>
</tr>
<tr>
<td>23.Web page 4 for Brand B</td>
<td>0</td>
<td>0</td>
<td>12</td>
<td>0</td>
</tr>
<tr>
<td>24.Web page 5 for Brand B</td>
<td>0</td>
<td>0</td>
<td>12</td>
<td>0</td>
</tr>
<tr>
<td>25.Web page 6 for Brand B: Attribute 1 and 2</td>
<td>0</td>
<td>0</td>
<td>51</td>
<td>0</td>
</tr>
<tr>
<td>26.Web page 1 for Brand D: Main</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>78</td>
</tr>
<tr>
<td>27.Web page 2 for Brand D: Attribute 4 basic</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>59</td>
</tr>
<tr>
<td>28.Web page 3 for Brand D: Attribute 4 level 3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>13</td>
</tr>
<tr>
<td>29.Web page 4 for Brand D: Attribute 4 level 1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>30.Web page 5 for Brand D: Attribute 4 level 2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>12</td>
</tr>
<tr>
<td>31.Word of mouth</td>
<td>87</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>32.Sticker talker</td>
<td>80</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 5.4 The frequency of media exposures
<table>
<thead>
<tr>
<th>The last shopping trip of pre-IA</th>
<th>The first shopping trip of post-IA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Buy Brand A+</td>
</tr>
<tr>
<td>Sticker and Buy Brand A</td>
<td>12</td>
</tr>
<tr>
<td>Sticker and Do not buy Brand A</td>
<td>17</td>
</tr>
</tbody>
</table>

Table 5.5 Sticker
Table 5.6 Number of respondents who do not select media of each brand

<table>
<thead>
<tr>
<th></th>
<th>Brand A+</th>
<th>Brand A</th>
<th>Brand B</th>
<th>Brand D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of respondents</td>
<td>53</td>
<td>48</td>
<td>38</td>
<td>42</td>
</tr>
</tbody>
</table>

*19 respondents do not select media of any brand.*
1. Baseline

<table>
<thead>
<tr>
<th>Model</th>
<th>Likelihood</th>
<th>Consumer Heterogeneity</th>
<th>Log marginal density</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Baseline</td>
<td>Equation (5.45) and (5.46)</td>
<td>$\tilde{g}_h^{(j)} \sim N\left(\widetilde{\mathbf{g}}^{(j)}, V^{(j)}_g\right)$, $j = 1, 2$</td>
<td>-4759.493</td>
</tr>
<tr>
<td>2. Two random-effect</td>
<td>Equation (5.17),(5.18), (5.47), and (5.48)</td>
<td>$\tilde{g}_h^{(1)} \sim N\left(\overline{\mathbf{g}}^{(1)}, V^{(1)}_g\right)$, $\tilde{g}_h^{(2)} \sim N\left(\overline{\mathbf{g}}^{(1)} + \lambda m_h, V^{(2)}_g\right)$</td>
<td>-5318.452</td>
</tr>
<tr>
<td>3. Multiplicative (standardized media exposure)</td>
<td>Equation (5.17),(5.18), (5.38), and (5.40)</td>
<td>$\tilde{g}_h^{(1)} \sim N\left(\overline{\mathbf{g}}, V_g\right)$, $\tilde{g}_h^{(2)} = C_h \tilde{g}_h^{(1)}$</td>
<td>-5116.341</td>
</tr>
<tr>
<td>4. Multiplicative (non-standardized media exposure)</td>
<td>Equation (5.17),(5.18), (5.38), and (5.40)</td>
<td>$\tilde{g}_h^{(1)} \sim N\left(\overline{\mathbf{g}}, V_g\right)$, $\tilde{g}_h^{(2)} = C_h \tilde{g}_h^{(1)}$</td>
<td>-5031.642</td>
</tr>
<tr>
<td>5. Additive</td>
<td>Equation (5.17),(5.18), (5.39), and (5.40)</td>
<td>$\tilde{g}_h^{(1)} \sim N\left(\overline{\mathbf{g}}, V_g\right)$, $\tilde{g}_h^{(2)} = \tilde{g}_h^{(1)} + \Delta_h$</td>
<td>-5492.606</td>
</tr>
</tbody>
</table>

Table 5.7 Model comparison
<table>
<thead>
<tr>
<th></th>
<th>Baseline Model</th>
<th>Two random effects</th>
<th>Multiplicative (standardized Media)</th>
<th>Multiplicative</th>
<th>Additive</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pre-IA</td>
<td>Post-IA</td>
<td>Pre-IA</td>
<td>Post-IA</td>
<td></td>
</tr>
<tr>
<td>Brand A+</td>
<td>-</td>
<td>-0.243 (0.026)</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Brand A</td>
<td>1.265 (0.069)</td>
<td>-0.166 (0.022)</td>
<td>0.506 (0.009)</td>
<td>0.647 (0.00715)</td>
<td>1.265 (0.069)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.563 (0.0073)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.612 (0.011)</td>
</tr>
<tr>
<td>Brand B</td>
<td>-</td>
<td>-</td>
<td>-0.070 (0.00703)</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-</td>
<td></td>
<td>-0.133 (0.0070)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.056 (0.019)</td>
</tr>
<tr>
<td>Brand C</td>
<td>-2.018 (0.042)</td>
<td>-5.551 (0.124)</td>
<td>-1.766 (0.011)</td>
<td>-1.326 (0.00411)</td>
<td>-2.018 (0.042)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-1.055 (0.0025)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-1.156 (0.006)</td>
</tr>
<tr>
<td>Brand D</td>
<td>-1.539 (0.057)</td>
<td>-1.920 (0.053)</td>
<td>-1.652 (0.024)</td>
<td>-0.730 (0.00651)</td>
<td>-1.539 (0.057)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.586 (0.0060)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.793 (0.008)</td>
</tr>
<tr>
<td>Brand E</td>
<td>-1.190 (0.041)</td>
<td>-2.679 (0.079)</td>
<td>-1.307 (0.014)</td>
<td>-0.591 (0.00902)</td>
<td>-1.190 (0.041)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.634 (0.0051)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.722 (0.005)</td>
</tr>
<tr>
<td>Deal</td>
<td>-0.028 (0.003)</td>
<td>-0.065 (0.002)</td>
<td>-0.005 (0.003)</td>
<td>-0.027 (0.00090)</td>
<td>-0.028 (0.003)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.001 (0.0006)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.072 (0.002)</td>
</tr>
<tr>
<td>Display</td>
<td>-0.392 (0.052)</td>
<td>0.054 (0.008)</td>
<td>0.090 (0.013)</td>
<td>-0.034 (0.00209)</td>
<td>-0.392 (0.052)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.014 (0.0023)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.028 (0.003)</td>
</tr>
<tr>
<td>Shelf talker</td>
<td>-0.795 (0.083)</td>
<td>-0.072 (0.012)</td>
<td>-0.092 (0.003)</td>
<td>-0.119 (0.00302)</td>
<td>-0.795 (0.083)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.083 (0.0013)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.012 (0.003)</td>
</tr>
<tr>
<td>Bonus pack</td>
<td>-0.269 (0.019)</td>
<td>-0.277 (0.008)</td>
<td>-1.071 (0.010)</td>
<td>0.022 (0.00202)</td>
<td>-0.269 (0.019)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.010 (0.0012)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.243 (0.010)</td>
</tr>
<tr>
<td>Attribute 1</td>
<td>-0.278 (0.005)</td>
<td>0.219 (0.005)</td>
<td>-0.192 (0.002)</td>
<td>-0.053 (0.00107)</td>
<td>-0.278 (0.005)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.030 (0.0006)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.052 (0.002)</td>
</tr>
<tr>
<td>Attribute 2</td>
<td>-0.161 (0.022)</td>
<td>-0.064 (0.002)</td>
<td>-0.539 (0.017)</td>
<td>-0.102 (0.00117)</td>
<td>-0.161 (0.022)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.020 (0.0008)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.032 (0.002)</td>
</tr>
<tr>
<td>Attribute 3</td>
<td>-0.777 (0.027)</td>
<td>-0.912 (0.027)</td>
<td>-1.169 (0.013)</td>
<td>-0.298 (0.00426)</td>
<td>-0.777 (0.027)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.044 (0.0010)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.361 (0.004)</td>
</tr>
<tr>
<td>Attribute 4-level 2</td>
<td>-0.137 (0.003)</td>
<td>-0.096 (0.001)</td>
<td>-0.166 (0.004)</td>
<td>-0.070 (0.00075)</td>
<td>-0.137 (0.003)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.004 (0.0006)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.142 (0.001)</td>
</tr>
<tr>
<td>Attribute 4-level 3</td>
<td>-0.310 (0.004)</td>
<td>-0.220 (0.002)</td>
<td>-0.350 (0.002)</td>
<td>-0.172 (0.00097)</td>
<td>-0.310 (0.004)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.079 (0.0009)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.318 (0.002)</td>
</tr>
<tr>
<td>Attribute 4-level 4</td>
<td>-0.149 (0.005)</td>
<td>-0.187 (0.04)</td>
<td>-0.102 (0.007)</td>
<td>-0.034 (0.00093)</td>
<td>-0.149 (0.005)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.014 (0.0023)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.081 (0.001)</td>
</tr>
<tr>
<td>ln(x+1)</td>
<td>0.225 (0.003)</td>
<td>0.247 (0.002)</td>
<td>0.295 (0.001)</td>
<td>0.345 (0.00088)</td>
<td>0.225 (0.003)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.310 (0.0008)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.737 (0.002)</td>
</tr>
<tr>
<td>ln(T-p)</td>
<td>-0.869 (0.126)</td>
<td>-0.895 (0.030)</td>
<td>-1.252 (0.004)</td>
<td>0.005 (0.00160)</td>
<td>-0.869 (0.126)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.093 (0.0023)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.550 (0.003)</td>
</tr>
</tbody>
</table>

Table 5.8 Posterior estimates of $\bar{\beta}$
<table>
<thead>
<tr>
<th></th>
<th>Baseline Model</th>
<th>Two random effects</th>
<th>Multiplicative (standardized Media)</th>
<th>Multiplicative</th>
<th>Additive</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pre-IA</td>
<td>Post-IA</td>
<td>Pre-IA</td>
<td>Post-IA</td>
<td></td>
</tr>
<tr>
<td>Brand A+</td>
<td>-</td>
<td>0.325 (0.178)</td>
<td>-</td>
<td>5.074 (0.397)</td>
<td>-</td>
</tr>
<tr>
<td>Brand A</td>
<td>11.686 (2.377)</td>
<td>0.203 (0.552)</td>
<td>9.139 (0.807)</td>
<td>9.871 (0.771)</td>
<td>1.854 (0.0538)</td>
</tr>
<tr>
<td>Brand B</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>2.958 (0.0476)</td>
</tr>
<tr>
<td>Brand C</td>
<td>6.689 (0.869)</td>
<td>0.633 (2.626)</td>
<td>4.533 (0.385)</td>
<td>10.586 (1.082)</td>
<td>1.327 (0.0128)</td>
</tr>
<tr>
<td>Brand D</td>
<td>14.785 (4.152)</td>
<td>0.175 (0.836)</td>
<td>11.589 (2.212)</td>
<td>9.669 (3.024)</td>
<td>2.270 (0.0278)</td>
</tr>
<tr>
<td>Brand E</td>
<td>10.291 (0.802)</td>
<td>0.206 (5.825)</td>
<td>6.273 (0.500)</td>
<td>21.432 (2.760)</td>
<td>2.274 (0.0356)</td>
</tr>
<tr>
<td>Deal</td>
<td>0.295 (0.002)</td>
<td>0.268 (0.001)</td>
<td>0.376 (0.003)</td>
<td>0.304 (0.002)</td>
<td>0.207 (0.0003)</td>
</tr>
<tr>
<td>Display</td>
<td>2.721 (0.134)</td>
<td>0.212 (0.011)</td>
<td>1.323 (0.110)</td>
<td>0.759 (0.022)</td>
<td>0.738 (0.0044)</td>
</tr>
<tr>
<td>Shelf talker</td>
<td>1.227 (0.026)</td>
<td>0.865 (0.008)</td>
<td>1.353 (0.036)</td>
<td>0.862 (0.028)</td>
<td>1.046 (0.0086)</td>
</tr>
<tr>
<td>Bonus pack</td>
<td>1.311 (0.068)</td>
<td>0.325 (0.008)</td>
<td>2.049 (0.198)</td>
<td>0.945 (0.032)</td>
<td>0.443 (0.0043)</td>
</tr>
<tr>
<td>Attribute 1</td>
<td>0.377 (0.003)</td>
<td>0.203 (0.002)</td>
<td>0.417 (0.003)</td>
<td>0.433 (0.004)</td>
<td>0.222 (0.0005)</td>
</tr>
<tr>
<td>Attribute 2</td>
<td>0.556 (0.020)</td>
<td>0.633 (0.001)</td>
<td>1.362 (0.093)</td>
<td>0.275 (0.002)</td>
<td>0.301 (0.0008)</td>
</tr>
<tr>
<td>Attribute 3</td>
<td>0.813 (0.040)</td>
<td>0.175 (0.011)</td>
<td>1.405 (0.063)</td>
<td>0.724 (0.010)</td>
<td>0.378 (0.0014)</td>
</tr>
<tr>
<td>Attribute 4-</td>
<td>0.229 (0.001)</td>
<td>0.206 (0.000)</td>
<td>0.304 (0.002)</td>
<td>0.240 (0.001)</td>
<td>0.202 (0.0005)</td>
</tr>
<tr>
<td>Attribute 4-</td>
<td>0.285 (0.002)</td>
<td>0.268 (0.001)</td>
<td>0.396 (0.003)</td>
<td>0.282 (0.001)</td>
<td>0.228 (0.0004)</td>
</tr>
<tr>
<td>Attribute 4-</td>
<td>0.318 (0.003)</td>
<td>0.212 (0.002)</td>
<td>0.381 (0.002)</td>
<td>0.423 (0.005)</td>
<td>0.216 (0.0004)</td>
</tr>
<tr>
<td>ln(x+1)</td>
<td>0.228 (0.001)</td>
<td>0.865 (0.001)</td>
<td>0.279 (0.001)</td>
<td>0.266 (0.001)</td>
<td>0.206 (0.0004)</td>
</tr>
<tr>
<td>ln(T-p)</td>
<td>0.944 (0.035)</td>
<td>0.325 (0.019)</td>
<td>1.107 (0.045)</td>
<td>0.926 (0.025)</td>
<td>0.439 (0.0013)</td>
</tr>
</tbody>
</table>

Table 5.9 Posterior mean of $V_{\beta}$
<table>
<thead>
<tr>
<th>Information sources</th>
<th>Multiplicative (standardized Media)</th>
<th>Multiplicative</th>
<th>Additive</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Medium</td>
<td>var</td>
<td>Medium</td>
</tr>
<tr>
<td>Television</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.Commercial-partial</td>
<td>0.152</td>
<td>0.676</td>
<td>0.669</td>
</tr>
<tr>
<td>2.Commercial-complete</td>
<td>0.151</td>
<td>0.577</td>
<td>-0.011</td>
</tr>
<tr>
<td>Magazine</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.Magazine</td>
<td>0.147</td>
<td>0.460</td>
<td>0.301</td>
</tr>
<tr>
<td>4.Coupon-page</td>
<td>0.135</td>
<td>0.434</td>
<td>0.069</td>
</tr>
<tr>
<td>DTC Brochure</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.Brochure-page1</td>
<td>-0.265</td>
<td>0.456</td>
<td>-0.100</td>
</tr>
<tr>
<td>6.Brochure-page2</td>
<td>0.172</td>
<td>0.478</td>
<td>-0.174</td>
</tr>
<tr>
<td>7.Brochure-page3</td>
<td>-0.088</td>
<td>0.363</td>
<td>-0.035</td>
</tr>
<tr>
<td>8.Brochure-coupon</td>
<td>0.149</td>
<td>0.323</td>
<td>-0.036</td>
</tr>
<tr>
<td>9.Brochure info-page1</td>
<td>0.206</td>
<td>0.314</td>
<td>-0.001</td>
</tr>
<tr>
<td>10.Brochure info-page2</td>
<td>0.237</td>
<td>0.331</td>
<td>-0.304</td>
</tr>
</tbody>
</table>

Table 5.10 Posterior estimates of $\bar{y}$
Table 5.10 Continued

<table>
<thead>
<tr>
<th>Information sources</th>
<th>Two random effects</th>
<th>Multiplicative (standardized Media)</th>
<th>Multiplicative</th>
<th>Additive</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Medium var</td>
<td>Medium var</td>
<td>Medium var</td>
<td>Medium var</td>
</tr>
<tr>
<td>Web pages</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. Web page 1 for Brand A+: Main</td>
<td>-0.014</td>
<td>0.258</td>
<td>-0.003</td>
<td>0.008</td>
</tr>
<tr>
<td>12. Web page 2 for Brand A+</td>
<td>-0.083</td>
<td>0.244</td>
<td>-0.297</td>
<td>0.011</td>
</tr>
<tr>
<td>13. Web page 3 for Brand A+</td>
<td>-0.278</td>
<td>0.243</td>
<td>-0.260</td>
<td>0.008</td>
</tr>
<tr>
<td>14. Web page 4 for Brand A+</td>
<td>-0.128</td>
<td>0.322</td>
<td>-0.040</td>
<td>0.015</td>
</tr>
<tr>
<td>15. Web page 5 for Brand A+: Attribute 1</td>
<td>0.103</td>
<td>0.186</td>
<td>-0.013</td>
<td>0.007</td>
</tr>
<tr>
<td>16. Web page 6 for Brand A+</td>
<td>-0.123</td>
<td>0.197</td>
<td>-0.049</td>
<td>0.009</td>
</tr>
<tr>
<td>17. Web page 7 for Brand A+</td>
<td>0.156</td>
<td>0.184</td>
<td>-0.050</td>
<td>0.007</td>
</tr>
<tr>
<td>18. Web page 1 for Brand A: Main</td>
<td>0.205</td>
<td>0.184</td>
<td>-0.032</td>
<td>0.011</td>
</tr>
<tr>
<td>19. Web page 2 for Brand A: Product</td>
<td>0.067</td>
<td>0.147</td>
<td>-0.075</td>
<td>0.007</td>
</tr>
<tr>
<td>20. Web page 1 for Brand B: Main</td>
<td>0.197</td>
<td>0.140</td>
<td>0.272</td>
<td>0.009</td>
</tr>
<tr>
<td>21. Web page 2 for Brand B</td>
<td>-0.037</td>
<td>0.140</td>
<td>-0.186</td>
<td>0.020</td>
</tr>
<tr>
<td>22. Web page 3 for Brand B: Product</td>
<td>0.207</td>
<td>0.133</td>
<td>0.135</td>
<td>0.014</td>
</tr>
<tr>
<td>23. Web page 4 for Brand B</td>
<td>-0.115</td>
<td>0.136</td>
<td>-0.179</td>
<td>0.014</td>
</tr>
<tr>
<td>24. Web page 5 for Brand B</td>
<td>0.166</td>
<td>0.190</td>
<td>-0.152</td>
<td>0.011</td>
</tr>
<tr>
<td>25. Web page 6 for Brand B: Attribute 1 and 2</td>
<td>-0.065</td>
<td>0.102</td>
<td>-0.022</td>
<td>0.009</td>
</tr>
<tr>
<td>26. Web page 1 for Brand D: Main</td>
<td>-0.069</td>
<td>0.091</td>
<td>0.271</td>
<td>0.010</td>
</tr>
<tr>
<td>27. Web page 2 for Brand D: Attribute 4-basic</td>
<td>0.209</td>
<td>0.083</td>
<td>-0.044</td>
<td>0.008</td>
</tr>
<tr>
<td>28. Web page 3 for Brand D: Attribute 4-level 3</td>
<td>-0.145</td>
<td>0.124</td>
<td>0.011</td>
<td>0.028</td>
</tr>
<tr>
<td>29. Web page 4 for Brand D: Attribute 4-level 1</td>
<td>0.334</td>
<td>0.082</td>
<td>0.066</td>
<td>0.009</td>
</tr>
<tr>
<td>30. Web page 5 for Brand D: Attribute 4-level 2</td>
<td>0.380</td>
<td>0.131</td>
<td>0.127</td>
<td>0.006</td>
</tr>
<tr>
<td>31. Word of mouth</td>
<td>0.059</td>
<td>0.075</td>
<td>-0.252</td>
<td>0.016</td>
</tr>
<tr>
<td>32. Sticker talker</td>
<td>0.129</td>
<td>0.064</td>
<td>-0.137</td>
<td>0.006</td>
</tr>
<tr>
<td>Information sources</td>
<td>Two random effects</td>
<td>Multiplicative (standardized Media)</td>
<td>Multiplicative</td>
<td>Additive</td>
</tr>
<tr>
<td>---------------------</td>
<td>--------------------</td>
<td>------------------------------------</td>
<td>----------------</td>
<td>---------</td>
</tr>
<tr>
<td></td>
<td>Medium</td>
<td>var</td>
<td>Medium</td>
<td>var</td>
</tr>
<tr>
<td>Television</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.Commercial-partial</td>
<td>2.937</td>
<td>2.676</td>
<td>1.296</td>
<td>0.032</td>
</tr>
<tr>
<td>2.Commercial-complete</td>
<td>2.682</td>
<td>2.232</td>
<td>1.143</td>
<td>0.016</td>
</tr>
<tr>
<td>Magazine</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.Magazine</td>
<td>2.318</td>
<td>1.923</td>
<td>1.164</td>
<td>0.023</td>
</tr>
<tr>
<td>4.Coupon-page</td>
<td>2.330</td>
<td>1.554</td>
<td>1.003</td>
<td>0.024</td>
</tr>
<tr>
<td>DTC Brochure</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.Brochure-page1</td>
<td>3.145</td>
<td>2.246</td>
<td>1.592</td>
<td>0.045</td>
</tr>
<tr>
<td>6.Brochure-page2</td>
<td>3.793</td>
<td>2.814</td>
<td>1.192</td>
<td>0.043</td>
</tr>
<tr>
<td>7.Brochure-page3</td>
<td>2.931</td>
<td>1.771</td>
<td>0.745</td>
<td>0.011</td>
</tr>
<tr>
<td>8.Brochure-coupon</td>
<td>2.850</td>
<td>1.379</td>
<td>1.469</td>
<td>0.052</td>
</tr>
<tr>
<td>9.Brochure info-page1</td>
<td>3.114</td>
<td>1.951</td>
<td>1.221</td>
<td>0.019</td>
</tr>
<tr>
<td>10.Brochure info-page2</td>
<td>3.872</td>
<td>2.533</td>
<td>0.999</td>
<td>0.015</td>
</tr>
</tbody>
</table>

Table 5.11 Posterior estimate of $V_\gamma$ (Continued)
### Table 5.11 Continued

<table>
<thead>
<tr>
<th>Information sources</th>
<th>Two random effects</th>
<th>Multiplicative (standardized Media)</th>
<th>Multiplicative</th>
<th>Additive</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Medium</td>
<td>var</td>
<td>Medium</td>
<td>var</td>
</tr>
<tr>
<td>Web pages</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. Web page 1 for Brand A+: Main</td>
<td>2.992</td>
<td>1.653</td>
<td>1.727</td>
<td>0.153</td>
</tr>
<tr>
<td>12. Web page 2 for Brand A+</td>
<td>3.373</td>
<td>2.300</td>
<td>1.134</td>
<td>0.016</td>
</tr>
<tr>
<td>13. Web page 3 for Brand A+</td>
<td>3.162</td>
<td>1.864</td>
<td>1.083</td>
<td>0.016</td>
</tr>
<tr>
<td>14. Web page 4 for Brand A+</td>
<td>3.445</td>
<td>1.989</td>
<td>1.324</td>
<td>0.032</td>
</tr>
<tr>
<td>15. Web page 5 for Brand A+: Attribute 1</td>
<td>3.041</td>
<td>1.795</td>
<td>1.251</td>
<td>0.057</td>
</tr>
<tr>
<td>16. Web page 6 for Brand A+</td>
<td>2.650</td>
<td>1.140</td>
<td>0.884</td>
<td>0.010</td>
</tr>
<tr>
<td>17. Web page 7 for Brand A+</td>
<td>3.042</td>
<td>1.690</td>
<td>1.123</td>
<td>0.025</td>
</tr>
<tr>
<td>18. Web page 1 for Brand A: Main</td>
<td>3.759</td>
<td>2.272</td>
<td>1.667</td>
<td>0.047</td>
</tr>
<tr>
<td>19. Web page 2 for Brand A: Products</td>
<td>2.618</td>
<td>1.198</td>
<td>1.053</td>
<td>0.019</td>
</tr>
<tr>
<td>20. Web page 1 for Brand B: Main</td>
<td>3.106</td>
<td>1.735</td>
<td>1.683</td>
<td>0.042</td>
</tr>
<tr>
<td>21. Web page 2 for Brand B</td>
<td>3.032</td>
<td>1.619</td>
<td>1.316</td>
<td>0.044</td>
</tr>
<tr>
<td>22. Web page 3 for Brand B: Product</td>
<td>3.355</td>
<td>1.612</td>
<td>1.668</td>
<td>0.122</td>
</tr>
<tr>
<td>23. Web page 4 for Brand B</td>
<td>4.072</td>
<td>2.834</td>
<td>1.114</td>
<td>0.017</td>
</tr>
<tr>
<td>24. Web page 5 for Brand B</td>
<td>4.229</td>
<td>3.115</td>
<td>1.385</td>
<td>0.035</td>
</tr>
<tr>
<td>25. Web page 6 for Brand B: Attribute 1 and 2</td>
<td>2.844</td>
<td>1.387</td>
<td>1.100</td>
<td>0.026</td>
</tr>
<tr>
<td>26. Web page 1 for Brand D: Main</td>
<td>2.762</td>
<td>1.193</td>
<td>1.419</td>
<td>0.029</td>
</tr>
<tr>
<td>27. Web page 2 for Brand D: Attribute 4-basic</td>
<td>2.998</td>
<td>1.553</td>
<td>1.173</td>
<td>0.052</td>
</tr>
<tr>
<td>28. Web page 3 for Brand D: Attribute 4-level 3</td>
<td>3.683</td>
<td>2.495</td>
<td>1.416</td>
<td>0.070</td>
</tr>
<tr>
<td>29. Web page 4 for Brand D: Attribute 4-level 1</td>
<td>2.918</td>
<td>1.571</td>
<td>1.496</td>
<td>0.025</td>
</tr>
<tr>
<td>30. Web page 5 for Brand D: Attribute 4-level 2</td>
<td>3.631</td>
<td>2.207</td>
<td>1.176</td>
<td>0.021</td>
</tr>
<tr>
<td>31. Word of mouth</td>
<td>2.970</td>
<td>1.395</td>
<td>1.148</td>
<td>0.022</td>
</tr>
<tr>
<td>32. Sticker talker</td>
<td>2.717</td>
<td>1.319</td>
<td>1.418</td>
<td>0.048</td>
</tr>
<tr>
<td>Source of Information</td>
<td>Multiplicative (Standardized Media)</td>
<td>Multiplicative</td>
<td>Additive</td>
<td></td>
</tr>
<tr>
<td>------------------------</td>
<td>-------------------------------------</td>
<td>----------------</td>
<td>----------</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Medium</td>
<td>Var</td>
<td>Medium</td>
<td>Var</td>
</tr>
<tr>
<td><strong>Television</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.Commercial-partial</td>
<td>0.165</td>
<td>0.009</td>
<td><strong>0.0602</strong></td>
<td>0.0013</td>
</tr>
<tr>
<td>2.Commercial-complete</td>
<td>0.091</td>
<td>0.007</td>
<td>-0.0073</td>
<td>0.0014</td>
</tr>
<tr>
<td><strong>Magazine</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.Magazine</td>
<td>0.113</td>
<td>0.008</td>
<td>0.0288</td>
<td>0.0012</td>
</tr>
<tr>
<td>4.Coupon-page</td>
<td>0.012</td>
<td>0.008</td>
<td>0.0173</td>
<td>0.0014</td>
</tr>
<tr>
<td><strong>Direct-to-Consumer Brochure</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.Brochure-page1</td>
<td>0.122</td>
<td>0.007</td>
<td>0.0075</td>
<td>0.0012</td>
</tr>
<tr>
<td>6.Brochure-page2</td>
<td>0.011</td>
<td>0.007</td>
<td>-0.0002</td>
<td>0.0012</td>
</tr>
<tr>
<td>7.Brochure-page3</td>
<td>0.018</td>
<td>0.007</td>
<td><strong>0.0634</strong></td>
<td>0.0012</td>
</tr>
<tr>
<td>8.Brochure-coupon</td>
<td>0.102</td>
<td>0.009</td>
<td>-0.0330</td>
<td>0.0014</td>
</tr>
<tr>
<td>9.Brochure info-page1</td>
<td>0.116</td>
<td>0.009</td>
<td>0.0315</td>
<td>0.0017</td>
</tr>
<tr>
<td>10.Brochure info-page2</td>
<td>0.058</td>
<td>0.009</td>
<td>-0.0014</td>
<td>0.0011</td>
</tr>
</tbody>
</table>

(Continued)

Table 5.12 Posterior estimates of $\bar{\theta}$
<table>
<thead>
<tr>
<th>Source of Information</th>
<th>Multiplicative (Standardized Media)</th>
<th>Multiplicative</th>
<th>Additive</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Medium</td>
<td>Var</td>
<td>Medium</td>
</tr>
<tr>
<td>Web pages</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. Web page 1 for Brand A+: Main</td>
<td>0.135</td>
<td>0.009</td>
<td>0.0399</td>
</tr>
<tr>
<td>12. Web page 2 for Brand A+</td>
<td>0.058</td>
<td>0.009</td>
<td>0.0191</td>
</tr>
<tr>
<td>13. Web page 3 for Brand A+</td>
<td>0.093</td>
<td>0.012</td>
<td>-0.0235</td>
</tr>
<tr>
<td>14. Web page 4 for Brand A+</td>
<td>0.072</td>
<td>0.008</td>
<td>-0.0050</td>
</tr>
<tr>
<td>15. Web page 5 for Brand A+: Attribute 1</td>
<td>0.117</td>
<td>0.007</td>
<td>-0.0455</td>
</tr>
<tr>
<td>16. Web page 6 for Brand A+</td>
<td>0.081</td>
<td>0.007</td>
<td>-0.0234</td>
</tr>
<tr>
<td>17. Web page 7 for Brand A+</td>
<td>0.075</td>
<td>0.007</td>
<td>-0.0054</td>
</tr>
<tr>
<td>18. Web page 1 for Brand A: Main</td>
<td>0.137</td>
<td>0.007</td>
<td>-0.0199</td>
</tr>
<tr>
<td>19. Web page 2 for Brand A: Products</td>
<td>0.109</td>
<td>0.009</td>
<td>0.0030</td>
</tr>
<tr>
<td>20. Web page 1 for Brand B: Main</td>
<td>0.031</td>
<td>0.007</td>
<td>-0.0046</td>
</tr>
<tr>
<td>21. Web page 2 for Brand B</td>
<td>0.054</td>
<td>0.006</td>
<td>-0.0102</td>
</tr>
<tr>
<td>22. Web page 3 for Brand B: Product</td>
<td>0.040</td>
<td>0.008</td>
<td>0.0136</td>
</tr>
<tr>
<td>23. Web page 4 for Brand B</td>
<td>0.132</td>
<td>0.006</td>
<td>-0.0255</td>
</tr>
<tr>
<td>24. Web page 5 for Brand B</td>
<td>0.181</td>
<td>0.008</td>
<td>-0.0095</td>
</tr>
<tr>
<td>25. Web page 6 for Brand B: Attribute 1 and 2</td>
<td>0.090</td>
<td>0.008</td>
<td>0.0126</td>
</tr>
<tr>
<td>26. Web page 1 for Brand D: Main</td>
<td>0.081</td>
<td>0.007</td>
<td>-0.0012</td>
</tr>
<tr>
<td>27. Web page 2 for Brand D: Attribute 4</td>
<td>0.060</td>
<td>0.007</td>
<td>0.0220</td>
</tr>
<tr>
<td>28. Web page 3 for Brand D: Attribute 4</td>
<td>-0.005</td>
<td>0.009</td>
<td>0.0063</td>
</tr>
<tr>
<td>29. Web page 4 for Brand D: Attribute 4</td>
<td>0.088</td>
<td>0.008</td>
<td>0.0272</td>
</tr>
<tr>
<td>30. Web page 5 for Brand D: Attribute 4</td>
<td>0.102</td>
<td>0.009</td>
<td>-0.0001</td>
</tr>
<tr>
<td>31. Word of mouth</td>
<td>0.110</td>
<td>0.007</td>
<td>-0.0419</td>
</tr>
<tr>
<td>32. Sticker talker</td>
<td>0.030</td>
<td>0.006</td>
<td>-0.0074</td>
</tr>
<tr>
<td>Source of Information</td>
<td>Multiplicative (Standardized Media)</td>
<td>Multiplicative</td>
<td>Additive</td>
</tr>
<tr>
<td>--------------------------------</td>
<td>-------------------------------------</td>
<td>----------------</td>
<td>----------</td>
</tr>
<tr>
<td></td>
<td>Medium</td>
<td>Var</td>
<td>Medium</td>
</tr>
<tr>
<td>Television</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Commercial-partial</td>
<td>25.567</td>
<td>0.376</td>
<td>4.190</td>
</tr>
<tr>
<td>2. Commercial-complete</td>
<td>26.158</td>
<td>0.572</td>
<td>4.200</td>
</tr>
<tr>
<td>Magazine</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Magazine</td>
<td>25.278</td>
<td>0.535</td>
<td>4.282</td>
</tr>
<tr>
<td>4. Coupon-page</td>
<td>25.899</td>
<td>0.370</td>
<td>4.371</td>
</tr>
<tr>
<td>Direct-to-Consumer Brochure</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Brochure-page1</td>
<td>26.333</td>
<td>0.437</td>
<td>4.376</td>
</tr>
<tr>
<td>6. Brochure-page2</td>
<td>25.954</td>
<td>0.681</td>
<td>4.344</td>
</tr>
<tr>
<td>7. Brochure-page3</td>
<td>26.632</td>
<td>0.460</td>
<td>4.400</td>
</tr>
<tr>
<td>8. Brochure-coupon</td>
<td>26.614</td>
<td>0.468</td>
<td>4.408</td>
</tr>
<tr>
<td>9. Brochure info-page1</td>
<td>26.631</td>
<td>0.306</td>
<td>4.414</td>
</tr>
<tr>
<td>10. Brochure info-page2</td>
<td>26.217</td>
<td>0.521</td>
<td>4.438</td>
</tr>
</tbody>
</table>

(Continued)

Table 5.13 Posterior estimates of $V_\theta$
<table>
<thead>
<tr>
<th>Source of Information</th>
<th>Multiplicative (Standardized Media)</th>
<th>Multiplicative</th>
<th>Additive</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Medium</td>
<td>Var</td>
<td>Medium</td>
</tr>
<tr>
<td>Web pages</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. Web page 1 for Brand A+: Main</td>
<td>26.216</td>
<td>0.691</td>
<td>4.435</td>
</tr>
<tr>
<td>12. Web page 2 for Brand A+</td>
<td>26.844</td>
<td>0.718</td>
<td>4.427</td>
</tr>
<tr>
<td>13. Web page 3 for Brand A+</td>
<td>26.745</td>
<td>0.662</td>
<td>4.370</td>
</tr>
<tr>
<td>14. Web page 4 for Brand A+</td>
<td>26.379</td>
<td>0.726</td>
<td>4.331</td>
</tr>
<tr>
<td>15. Web page 5 for Brand A+: Attribute 1</td>
<td>26.660</td>
<td>0.420</td>
<td>4.426</td>
</tr>
<tr>
<td>16. Web page 6 for Brand A+</td>
<td>26.633</td>
<td>0.385</td>
<td>4.443</td>
</tr>
<tr>
<td>17. Web page 7 for Brand A+</td>
<td>26.720</td>
<td>0.411</td>
<td>4.379</td>
</tr>
<tr>
<td>18. Web page 1 for Brand A: Main</td>
<td>26.540</td>
<td>0.569</td>
<td>4.418</td>
</tr>
<tr>
<td>19. Web page 2 for Brand A: Products</td>
<td>26.974</td>
<td>0.389</td>
<td>4.362</td>
</tr>
<tr>
<td>20. Web page 3 for Brand B: Main</td>
<td>25.967</td>
<td>0.758</td>
<td>4.462</td>
</tr>
<tr>
<td>21. Web page 2 for Brand B</td>
<td>26.295</td>
<td>0.470</td>
<td>4.456</td>
</tr>
<tr>
<td>22. Web page 3 for Brand B: Product</td>
<td>26.232</td>
<td>0.483</td>
<td>4.351</td>
</tr>
<tr>
<td>23. Web page 4 for Brand B</td>
<td>26.746</td>
<td>0.426</td>
<td>4.355</td>
</tr>
<tr>
<td>24. Web page 5 for Brand B</td>
<td>26.324</td>
<td>0.492</td>
<td>4.351</td>
</tr>
<tr>
<td>25. Web page 6 for Brand B: Attribute 1 and 2</td>
<td>26.798</td>
<td>0.472</td>
<td>4.397</td>
</tr>
<tr>
<td>26. Web page 1 for Brand D: Main</td>
<td>26.753</td>
<td>0.361</td>
<td>4.420</td>
</tr>
<tr>
<td>27. Web page 2 for Brand D: Attribute 4</td>
<td>26.128</td>
<td>0.445</td>
<td>4.521</td>
</tr>
<tr>
<td>28. Web page 3 for Brand D: Attribute 4</td>
<td>26.574</td>
<td>0.348</td>
<td>4.478</td>
</tr>
<tr>
<td>29. Web page 4 for Brand D: Attribute 4</td>
<td>27.022</td>
<td>0.400</td>
<td>4.422</td>
</tr>
<tr>
<td>30. Web page 5 for Brand D: Attribute 4</td>
<td>26.376</td>
<td>0.380</td>
<td>4.299</td>
</tr>
<tr>
<td>31. Word of mouth</td>
<td>26.244</td>
<td>0.372</td>
<td>4.331</td>
</tr>
<tr>
<td>32. Sticker talker</td>
<td>26.141</td>
<td>0.512</td>
<td>4.367</td>
</tr>
</tbody>
</table>
CHAPTER 6

CONCLUSIONS

Consumer decisions are driven by latent behavioral processes that are heterogeneous and non-linear. Latent variable models are an important aspect of consumer research whenever the determinants of behavior are an important aspect of study. Complexities of studying consumer decisions is that factors that affect latent behavior processes are often not observed by researchers. For example, a consumer’s brand preferences are affected by various information sources and vary over time. A consumer’s purchase timing decision is affected by a latent inventory process. When the inventory of a product goes low, a consumer needs to restock the inventory.

The purpose of this thesis is to develop methods of dealing with heterogeneous, non-linear models of behavior for two issues commonly encountered in marketing. The first issue relates to a consumer’s purchase decision driven by a time-varying latent behavior process. The second issue relates to consumer preferences affected by multiple information sources.

This thesis comprises three essays. The first issue is addressed in the first and the second essays, and the second issue is addressed in the third essay. The first essay develops a new method of error augmentation for state-space models of economic
behavior where the observed behavior is related to a latent variable whose temporal
temporal variation is described by a state equation. A new method of data augmentation for
evaluating the likelihood is needed because of the presence of common parameters in the
observation and state equations, and shared error realizations.

The proposed state-space model is applied to analyze a consumer’s purchase and
resignation decisions in a membership club. The result shows that the relatively short
longevity of many customers is due to the high autocorrelation of the state variable,
where customer inventory is not depleted sufficiently fast in light of the frequency of new
offers arriving by mail. Increasing inter-arrival time between shipments can lead to
longer customer longevity and greater sales.

The new method of error augmentation developed in the first essay is applied to the
second and the third essays in this thesis. The standard data augmentation cannot be
applied to models proposed in this thesis because the observed discrete choices do not
have direct correspondence to the errors. As a result, the errors cannot be generated
directly form a distribution, and the likelihood functions of proposed models are difficult
to compute without the new approach. The second chapter of this thesis provides an
exclusive discussion of data augmentation. An example of a binary probit model is
provided to compare the conceptual difference between data augmentation, error
augmentation, and the proposed method of error augmentation.

The second essay investigates an alternative method of modeling customer inter-
purchase times. In traditional models in direct marketing, inter-purchase times are treated
as dependent variables whose model parameters are used to identify profitable customers.
In this essay, a state-space model is proposed to investigate the possibility to model inter-
purchase times as an independent variable that is affected by a time-varying latent inventory process.

The proposed model is illustrated by a data set of business-to-business selling, and a data set of business-to-customer selling. The results indicate that the proposed state-space model can accurately describe customer behavior when the specification of the state equation is plausible for the data. Purchases in the business-to-business dataset are more regular than those in the business-to-consumer dataset, most likely because business decisions are made for longer period of time. The proposed state-space model that assumes a customer’s purchase decision is triggered by a latent inventory system offers a plausible description of purchases decisions in the business-to-business dataset.

In the third essay, a demand model is proposed to address three issues in choice modeling. The first issue relates to the effects of multiple treatments for data collected in a pre-post study. Since pre-post measurements are a pair of observations of a respondent, the effects of multiple treatments are studied by relating pre and post measurements to common random-effect realizations, and by specifying an explicit relationship between pre and post measurements.

The second issue relates to line extension that is widely adopted in marketing practice. In this essay, a consumer’s preference to a new brand is decomposed into three sources: media effect of new brand, line extension effect, and other unobserved factor. This approach allows researchers to diagnose the source that gives most impacts on consumer preferences to a new brand.

The last issue relates to consumer decisions of brand-pack and no-choice for consumer packaged goods at the level of stock-keeping unit. Translated Cobb-Douglas
utility is used to accommodate no-choice decisions. A direct evaluation procedure and a new variant of data augmentation are used to overcome estimation difficulties associated with no-choice decisions.

Data from a leading packaged goods company are used to illustrate the method by investigating changes in consumer preferences and sensitivities in a simulated shopping environment. The results indicate that consumers’ reactions to media are very heterogeneous. Media can make some respondents’ posttest consumer preferences go to extreme, and make some respondents have more homogeneous consumer preferences. Despite extreme consumer preferences and sensitivities are observed, between-brand variation becomes smaller after respondents review media.

This thesis contributes marketing literature by developing a new method of error augmentation for latent variable models that cannot be estimated by standard approaches. The new method of error augmentation is illustrated by three different marketing applications in this thesis. The state-space model proposed in the first and the second essays can be extended to study consumer learning or consumer searching behavior. The demand model proposed in the third essay can be extended to study consumer preference changes in multiple stages.

The proposed estimation algorithm and the new method of error augmentation suffer from slow convergence of Markov chain. It would be interesting to see whether the convergence of Markov chain can be improved by more efficient algorithm.
APPENDIX A

MCMC ESTIMATION FOR ESSAY 1
Estimation of the parameters of the model described by Equations (3.35)-(3.37), with parameters and prior distributions described by Equations (3.38)-(3.42) proceeds by recursively generating from the following conditional distributions:

(1) Generate \( \{s_{t,j}, t = 1, 2, \ldots, T_j; j = 1, 2, \ldots, H\} \)

The draw of \( s_{t,j} \) is generated from the conditional normal distribution specified by the joint distribution of the vector \( S_j = \{s_{1,j}, s_{2,j}, \ldots, s_{T_j,j}\} \) and the truncation points \( T_{y_{i,j}} \) and \( T_{x_{i,j}} \) determined by the decision rules in both observation equations.

The joint distribution of \( S_j = \{s_{1,j}, s_{2,j}, \ldots, s_{T_j,j}\} \) is

\[
S_j \sim N(\mu_j, \Sigma_j)
\]

where

\[
\mu_j = \begin{bmatrix}
\mu_{s_{1,j}} = \phi_j s_{u_{1,j}} + \exp(\beta_j) \\
\mu_{s_{2,j}} = \phi_j \mu_{s_{1,j}} + \exp(\alpha_{s(1)}^* + \beta_j^*) y_{i,j} \\
\mu_{s_{3,j}} = \phi_j \mu_{s_{2,j}} + \exp(\alpha_{s(2)}^* + \beta_j^*) y_{2,j} \\
\mu_{s_{4,j}} = \phi_j \mu_{s_{3,j}} + \exp(\alpha_{s(3)}^* + \beta_j^*) y_{3,j} \\
\vdots \\
\mu_{s_{T_j,j}} = \phi_j \mu_{s_{T_j-1,j}} + \exp(\alpha_{s(T-1)}^* + \beta_j^*) y_{T_j-1,j}
\end{bmatrix}
\]

and

\[
\Sigma_j = \begin{bmatrix}
1 & \phi_j & \phi_j^2 & \phi_j^3 & \cdots & \phi_j^{T-1} \\
\phi_j & \phi_j^2 + 1 & \phi_j(\phi_j^2 + 1) & \phi_j^2(\phi_j^2 + 1) & \cdots & \phi_j^{T-2}(\phi_j^2 + 1) \\
\phi_j^2 & \phi_j(\phi_j^2 + 1) & \phi_j^4 + \phi_j^2 + 1 & \phi_j^2(\phi_j^4 + \phi_j^2 + 1) & \cdots & \vdots \\
\phi_j^3 & \phi_j^2(\phi_j^2 + 1) & \phi_j^4 + \phi_j^2 + 1 & \phi_j^4 + \phi_j^2 + 1 & \cdots & \vdots \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
\phi_j^{T-1} & \phi_j^{T-2}(\phi_j^2 + 1) & \phi_j^{T-3}(\phi_j^4 + \phi_j^2 + 1) & \cdots & \phi_j^{(T-2)^2} + 1 & \phi_j^{(T-1)^2} + \phi_j^{(T-2)^2} + \cdots + \phi_j^2 + 1
\end{bmatrix}
\]
Let $\tau_{ii}$ be the $i^{th}$ element of $(1/diag(\Sigma_j^{-1}))$, $\Sigma_{ij,ij}$ be the $t^{th}$ column of covariance matrix $\Sigma_j$ excluding the $t^{th}$ element from this vector, and $\Sigma_{tj,-tj}$ be a matrix after removing the $t^{th}$ column and the $t^{th}$ vector from the covariance matrix $\Sigma_j$. Then, $s_{t,j}$ can be generated from the full conditional

$$[s_{t,j} | s_{-t,j}, y_{t:j}] \propto N[\mu_{s_{t,j}} + \Sigma'_{g,-g} \Sigma_{-g,-g}^{-1} (s_{-t,j} - \mu_{s_{-t,j}}), \tau_{tt,j}] \cdot \Phi(y_{t,j}, x_{t,j})$$  \quad (A2)$$

Replacing $s_{t,j}$ with $Ty_{t,j}$ in the decision rule of purchase, the truncation point associated with the purchase decision can be derived by solving $Ty_{t,j}$ in the following function with a Newton-Raphson Algorithm.

$$[Ty_{t,j} + \exp(\alpha^*_k + \beta^*_j)]^{\rho} - Ty_{t,j}^{\rho} = \gamma$$  \quad (A3)$$

Replacing $s_{t,j}$ with $Tx_{t,j}$ in the decision rule of resigning, the truncation point can be computed by

$$Tx_{t,j} = \left(\frac{\delta}{\rho}\right)^{1/\rho} - \exp(\alpha^*_k + \beta^*_j) \cdot y_{t,j}$$  \quad (A4)$$

The observed decisions of a household $j$ at time $t$ can have four different outcomes: purchase and stay ($y_{t,j}=1, x_{t,j}=0$), purchase and resign ($y_{t,j}=1, x_{t,j}=1$), return and stay ($y_{t,j}=0, x_{t,j}=0$), and return and resign ($y_{t,j}=0, x_{t,j}=1$). The inverse CDF method is used to generate $s_{t,j}$ from univariate normal distribution specified in Equation (A2) with the truncation points corresponding to the observed data shown below.
The remaining parameters are generated using a random-walk Metropolis-Hastings algorithm (see Chib and Greenberg 1995), where error term realizations are constructed from the draws of \( \{s_{t,j}\} \), and the candidate parameter values are used to reconstruct values of the state variables.

(2) Generate \( \{\beta_j^*, j = 1, 2, \cdots, H\} \)

The stochastic terms are obtained by

\[
\begin{align*}
\varepsilon_{1,j} &= s_{1,j} - s_{1,j} \cdot \exp(\beta_j^*) \\
\varepsilon_{t,j} &= s_{t,j} - s_{t-1,j} \cdot \exp(\alpha_{k(t-1)}^* + \beta_j^*) \cdot y_{t-1,j}
\end{align*}
\]  
(A5)

Let \( \beta_j^{*(p)} \) be the previous draw for \( \beta_j^* \), and the next draw \( \beta_j^{*(n)} \) is given by

\[
\beta_j^{*(n)} = \beta_j^{*(p)} + \Delta \beta,
\]

where \( \Delta \beta \) is a draw from a candidate generating density Normal (0, 0.1). The choice for parameters of this density ensures 30% acceptance rate. Let

\[
\omega_0 = \exp(\beta_j^{*(n)})
\]

\[
\omega_t = \exp(\alpha_{k(t)}^* + \beta_j^{*(n)}), \text{ for } t = 1, \cdots, T_j
\]

And, compute

\[
\begin{align*}
s_1 &= \phi_1 s_{0,j} + \omega_0 + \varepsilon_{1,j} \\
s_t &= \phi_t s_{t-1,j} + \omega_t \cdot y_{t-1,j} + \varepsilon_{t,j}, \text{ for } t = 1, \cdots, T_j
\end{align*}
\]
Compute $\omega_{0,j}^{(p)}$ and $\omega_{t,j}^{(p)}$ from the same algebra with the previous draws of $\beta_{*j}^*$, and let $s_{1,j}^{(p)}$ and $s_{t,j}^{(p)}$ be the previous draw of $s_{1,j}$ and $s_{t,j}$. Thus, the probability of move $\beta_{*j}^*$ is given by

$$\kappa(\beta_{j}^{(p)}, \beta_{j}^{(n)}) = \min \left( \frac{\pi(\beta_{j}^{(n)})}{\pi(\beta_{j}^{(p)}),1} \right)$$

and

$$\frac{\pi(\beta_{j}^{(n)})}{\pi(\beta_{j}^{(p)})} = \exp \left( \left( -\frac{1}{2} (\epsilon^T \Sigma_{\epsilon}^{-1} \epsilon) + \frac{1}{2} \epsilon \Sigma_{\epsilon}^{-1} \epsilon + \frac{1}{2} (\epsilon^T \Sigma_{\epsilon}^{-1} \epsilon) \right) \right) \frac{\prod_{i=1}^{t} l(y_{i,j}, x_{i,j} | \beta_{j}^{(n)})}{\prod_{i=1}^{t} l(y_{i,j}, x_{i,j} | \beta_{j}^{(p)})}$$

(3) Generate $\{\alpha_{k}^*, k = 1, 2, \ldots, K\}$

The algorithm to generate $\alpha_{k}^*$ is different from the algorithm to generate household-specified parameters such as $\beta_{*j}^*$. The draw of $\alpha_{k}^*$ will affect not only the purchase decision to the item $k$ at the time $t$ period but also both decisions after the time period $t$. In additions, different household may receive the same item at different time period. Thus, the candidate of $\alpha_{k}^*$ needed to be evaluated across households and time periods.
For item $k$, obtain the stochastic terms by Equation (A3). Let $\alpha_{k}^{*(p)}$ be the previous draw for $\alpha_{k}^{*}$, and the next draw $\alpha_{k}^{*(n)}$ is given by

$$\alpha_{k}^{*(n)} = \alpha_{k}^{*(p)} + \Delta \alpha,$$

where $\Delta \alpha$ is a draw from a candidate generating density Normal (0, 0.008). The choice for parameters of this density ensures 30% acceptance rate. Let

$$\omega_{t,j}^{(n)} = \exp(\alpha_{k(t)}^{*(n)} + \beta_{j}^{*}), \text{ for } t = 1, \cdots, T_{j}$$

And $s_{t,j}^{(n)}$ is computed by

$$s_{t,j}^{(n)} = \phi_{j} s_{t-1,j}^{(n)} + \omega_{t-1,j}^{(n)} \gamma_{t-1,j} + \epsilon_{t,j}, \text{ for } t = 2, \cdots, T_{j} \text{ and } j = 1, 2, \cdots, H$$

Compute $\omega_{t,j}^{(p)}$ from the same algebra with the previous draws of $\alpha_{k}^{*}$, and let $s_{t,j}^{(p)}$, and $s_{t,j}^{(p)}$ be the previous draws of $s_{t,j}$ and $s_{t,j}$.

Thus, the probability of move $\alpha_{k}^{*}$ is given by

$$\kappa(\alpha_{k}^{*(p)}, \alpha_{k}^{*(n)}) = \min \left( \frac{\pi(\alpha_{k}^{*(n)})}{\pi(\alpha_{k}^{*(p)})}, 1 \right)$$

and

161
where $t(k)$ denotes the time period that a household $j$ receives the item $k$.

(4) Generate $\{\phi_j^*, j = 1, 2, \ldots, H\}$

Obtain the stochastic terms by Equation (A3), and let $\phi_j^{*(p)}$ be the previous draw for $\phi_j^*$, and the next draw $\phi_j^{*(n)}$ is given by

$$
\phi_j^{*(n)} = \phi_j^{*(p)} + \Delta \phi ,
$$

where $\Delta \phi$ is a draw from a candidate generating density Normal (0, 2.0). The choice for parameters of this density ensures 30% acceptance rate. Let

$$
\omega_{0,j} = \exp(\beta_j^*)
$$
$$
\omega_{t,j} = \exp(\alpha_{k(t)}^* + \beta_j^*), \text{ for } t = 1, \ldots, T_j
$$
$$
\phi_j^{(n)} = \frac{\exp(\phi_j^{*(n)})}{1 + \exp(\phi_j^{*(n)})}
$$

and compute
\[ s_j^{(n)} = \phi_j^{(n)} s_{0,j} + \omega_{0,j} + \epsilon_{1,j} \]
\[ s_j^{(n)} = \phi_j^{(n)} s_{1-j} + \omega_{1-j} + \epsilon_{t-j}, \quad \text{for } t = 1, \cdots, T_j \]

Let \( s_{1,j}^{(p)} \) and \( s_{0,j}^{(p)} \) be the previous draw of \( s_{1,j} \) and \( s_{0,j} \). Thus, the probability of move \( \phi_j^* \) is given by

\[
\kappa(\phi_j^{*(p)}, \phi_j^{*(n)}) = \min \left( \frac{\pi(\phi_j^{*(n)})}{\pi(\phi_j^{*(p)})}, 1 \right)
\]

and

\[
\pi(\phi_j^{(n)}) = \frac{\exp \left( -\frac{1}{2} \sum_{j=0}^{2n-1} (\phi_j^{(n)} - \mu_j)^2 \right) \exp \left( -\frac{1}{2\sigma_j^2} (\phi_j^{(n)} - \overline{\phi_j}) \right) \prod_{j} p(y_{ij} | x_{ij} | \phi_j^{(n)})}{\exp \left( -\frac{1}{2} \sum_{j=0}^{2n-1} (\phi_j^{(n)} - \mu_j)^2 \right) \exp \left( -\frac{1}{2\sigma_j^2} (\phi_j^{(n)} - \overline{\phi_j}) \right) \prod_{j} p(y_{ij} | x_{ij} | \phi_j^{(n)})}
\]

(5) Generate \( \{s_{0,j}^*, j = 1, 2, \cdots, H\} \)

The stochastic terms are obtained by Equation (A3). Let \( s_{0,j}^{*(p)} \) be the previous draw for \( s_{0,j}^* \), and the next draw \( s_{0,j}^{*(n)} \) is given by

\[
s_{0,j}^{*(n)} = s_{0,j}^{*(p)} + \Delta s_{0,j},
\]

where \( \Delta s_{0,j} \) is a draw from a candidate generating density Normal (0, 1.2). The choice for parameters of this density ensures more than 30% acceptance rate. Let
\[ s_{0,j} = \exp(s_{0,j}^*) \]
\[ \omega_{0,j} = \exp(\beta_j^*) \]
\[ \omega_{t,j} = \exp(\alpha_{k(t)}^* + \beta_j^*), \text{ for } t = 1, \cdots, T_j \]

and compute

\[ s_{t,j}^{(n)} = \phi_j s_{0,j}^{(n)} + \omega_{0,j} + \epsilon_{t,j} \]
\[ s_{t,j}^{(n)} = \phi_j s_{0,j}^{(n)} + \omega_{t-1,j} \cdot y_{t-1,j} + \epsilon_{t,j}, \text{ for } t = 2, \cdots, T_j \]

Let \( s_{1,j}^{(p)} \) and \( s_{0,j}^{(p)} \) be the previous draw of \( s_{1,j} \) and \( s_{0,j} \). Thus, the probability of move \( s_{0,j}^* \) is given by

\[ \kappa(s_{0,j}^{*}, s_{0,j}^{*}) = \min \left( \frac{\pi(s_{0,j}^{*})}{\pi(s_{0,j}^{*})}, 1 \right) \]

and

\[
\frac{\pi(s_{0,j}^{(p)})}{\pi(s_{0,j}^{(p)})} = \exp \left( -\frac{1}{2} \left[ \begin{array}{c}
    \sum_{j}^{n} - \mu_j^n \\
    \vdots \\
    \sum_{j}^{n} - \mu_j^n
  \end{array} \right] \right) \exp \left( -\frac{1}{2\sigma^2} (s_{0,j}^{(p)} - \tilde{y}) \right) \prod_{t=1}^T p(y_{t,j}|x_{j,t} | s_{0,j}^{(p)})
\]

\[
= \exp \left( -\frac{1}{2} \left[ \begin{array}{c}
    \sum_{j}^{n} - \mu_j^n \\
    \vdots \\
    \sum_{j}^{n} - \mu_j^n
  \end{array} \right] \right) \exp \left( -\frac{1}{2\sigma^2} (s_{0,j}^{(p)} - \tilde{y}) \right) \prod_{t=1}^T p(y_{t,j}|x_{j,t} | s_{0,j}^{(p)})
\]
(6) Generate $\sigma^2_\beta$

Assume the prior distribution $[\sigma^2_\beta] \sim IG(f_\beta, g_\beta)$, $\sigma^2_\beta$ can be generated from the full conditional

$$
[\sigma^2_\beta | \text{rest}] = \prod_{j=1}^{H} [\beta^*_j | \overline{\beta}, \sigma^2_\beta][\sigma^2_\beta | f_\beta, g_\beta]
$$

$$
\propto IG \left( \frac{H}{2} + f_\beta, \left[ \frac{1}{2} \sum_{j=1}^{H} (\beta^*_j - \overline{\beta})^2 + \frac{1}{g_\beta} \right]^{-1} \right)
$$

(7) Generate $\overline{\phi}$

Assume the prior distribution $[\phi] \sim N(a_\phi, b^2_\phi)$ and $[\overline{\phi}] \sim N(a_\phi, b^2_\phi)$. $\overline{\phi}$ can be generated from the full conditional

$$
[\overline{\phi} | \text{rest}] \propto N \left( \frac{a_\phi / b^2_\phi + \sum_{j=1}^{H} \phi_j^* / \sigma^2_\phi}{1/b^2_\phi + H / \sigma^2_\phi}, \frac{1}{1/b^2_\phi + H / \sigma^2_\phi} \right)
$$

(8) Generate $\sigma^2_\phi$

Assume the prior distribution $[\sigma^2_\phi] \sim IG(f_\phi, g_\phi)$. $\sigma^2_\phi$ can be generated from the full conditional

$$
[\sigma^2_\phi | \text{rest}] = \prod_{j=1}^{H} [\phi^*_j | \overline{\phi}, \sigma^2_\phi][\sigma^2_\phi | f_\phi, g_\phi]
$$

$$
\propto IG \left( \frac{H}{2} + f_\phi, \left[ \frac{1}{2} \sum_{j=1}^{H} (\phi^*_j - \overline{\phi})^2 + \frac{1}{g_\phi} \right]^{-1} \right)
$$
APPENDIX B

MCMC ESTIMATION FOR ESSAY 2
B.1 State Space Model

The variant of error augmentation method is used to estimate the state-space model. Let \( j \) be the \( j \text{th} \) customer, \( t \) be the time period of purchase, \( s_{t,j} \) be the latent inventory of customer \( j \) at the time period \( t \), and \( w_{t,j} \) be the amount purchased by customer \( j \) at time period \( t \).

(1) Generate \( \{s_{i,j}, t = 2, \ldots, T_j; j = 1, 2, \ldots, J\} \)

The draw of \( s_{t,j} \) is generated from the conditional normal distribution specified by the joint distribution of the vector \( s_j = \{s_{2,j}, \ldots, s_{T_j,j}\} \). The joint distribution of \( s_j = \{s_{2,j}, \ldots, s_{T_j,j}\} \) is

\[
 s_j \sim N(\mu_j, \Sigma_j) \tag{B1}
\]

Where

\[
 \mu_j = \begin{bmatrix}
 -\theta_j + w_{1,j} y_{1,j} \\
 -2\theta_j + (w_{1,j} y_{1,j} + w_{2,j} y_{2,j}) \\
 \vdots \\
 -(T_j - 1)\theta_j + \sum_{i=1}^{T_j-1} w_{i,j} y_{i,j}
 \end{bmatrix}
 \text{ and } \Omega_j = \sigma_j^2 \Sigma_j = \sigma_j^2 \\
= \begin{bmatrix}
 1 & 1 & \cdots & 1 \\
 1 & 2 & \cdots & 2 \\
 \vdots & \vdots & \ddots & \vdots \\
 1 & 2 & \cdots & T_j - 1
 \end{bmatrix}
\]

Let \( \tau_{ii} \) be the \( i \text{th} \) element of \((1/\text{diag}(\Omega_j^{-1}))\), \( \Omega_{j,-t} \) be the \( t \text{th} \) column of covariance matrix \( \Omega_j \) excluding the \( t \text{th} \) element from this vector, and \( \Omega_{-t,-t} \) be a matrix after removing the \( t \text{th} \) column and the \( t \text{th} \) vector from the covariance matrix \( \Omega_j \). Then, \( s_{t,j} \) can be generated from the full conditional distribution:

\[
 [s_{t,j}, s_{-,t,j}, \text{rest}] \propto N[\mu_{s_{t,j}} + \Omega_{j,-t} \Omega_{j,-t}^{-1} (s_{t,j} - \mu_{s_{t,j}}), \tau_{s_{t,j}}] \cdot I(y_{t,j}) \tag{B2}
\]

167
(2) Generate \( \{ \theta^*_j, j = 1, 2, \ldots, J \} \)

A random-walk Metropolis-Hasting algorithm is used to generate draws of \( \theta^*_j \)'s (Tanner and Wong 1987, Chib and Greenberg 1995). The stochastic terms are obtained by

\[
e_{t,j} = s_{t,j} - s_{t-1,j} + \theta^*_j - w_{t-1,j}, \forall t = 2, \ldots, T_j
\]

Let \( \theta^*_{j}^{(p)} \) be the previous draw for \( \theta^*_j \), and the next draw \( \theta^*_{j}^{(n)} \) is given by

\[
\theta^*_{j}^{(n)} = \theta^*_{j}^{(p)} + \Delta \theta,
\]

where \( \Delta \theta \) is a draw from a candidate generating density Normal (0, 0.03) for the business-to-business dataset and a candidate generating density Normal (0, 2.0) for the business-to-customer dataset. The choice for parameters of this density ensures 30% acceptance rate.

Let \( s_{(n)}_{i,j} \) be the previous state variable, and the new value of state variable \( s_{(n)}_{i,j} \) given the candidate draw \( \theta^*_{j}^{(n)} \) can be computed by

\[
s_{(n)}_{i,j} = s_{(n)}_{i-1,j} - \theta^*_{j}^{(n)} + w_{i-1,j}, \forall t = 2, \ldots, T_j
\]

Thus, the probability of move \( \theta^*_j \) is given by

\[
f(\theta^*_j^{(p)}, \theta^*_j^{(n)}) = \min \left( \frac{\pi(\theta^*_j^{(n)})}{\pi(\theta^*_j^{(p)})}, 1 \right)
\]
\[
\pi(\theta_j^{(n)}) = \frac{\exp\left(-\frac{1}{2}(s_j^{(n)} - \mu_j)^\top \Omega_j (s_j^{(n)} - \mu_j)\right)}{\pi(\theta_j^{(p)})} = \frac{\exp\left(-\frac{1}{2\sigma^2} (\theta_j^{(n)} - \bar{\theta})^2\right) \prod_{i=2}^{T_j} I(y_{i,j} | s_{i,j}^{(n)})}{\exp\left(-\frac{1}{2\sigma^2} (\theta_j^{(p)} - \bar{\theta})^2\right) \prod_{i=2}^{T_j} I(y_{i,j} | s_{i,j}^{(p)})},
\]

where \( e_{j} = (e_{2,j}, e_{3,j}, \cdots, e_{T_j,j}) \)

(3) Generate \( \bar{\theta} \)

Assume the prior distribution \( \bar{\theta} \sim N(a, b^2) \). \( \bar{\theta} \) can be generated from

the full conditional

\[
[\bar{\theta} | \text{rest}] \sim N \left( \begin{bmatrix} J \frac{1}{\sigma^2} + \frac{1}{b^2} \end{bmatrix}^{-1} \left[ \frac{\sum_{j=1}^J \theta_j}{\sigma^2} + \frac{a}{b^2} \right], \left[ \frac{J}{\sigma^2} + \frac{1}{b^2} \right]^{-1} \right)
\]
(4) Generate $\sigma^2_{\theta}$

Assume the prior distribution $\sigma^2_{\theta} \sim IG(f_{\theta}, g_{\theta})$. $\sigma^2_{\theta}$ can be generated from the full conditional

$$\left[ \sigma^2_{\theta} \mid \text{rest} \right] \propto IG\left(\frac{J}{2} + f_{\theta}, \left[ \frac{1}{2} \sum_{j=1}^{J} (\theta_j - \bar{\theta})^2 + \frac{1}{g_{\theta}} \right]^{-1}\right)$$

(5) Generate $\{\sigma^2_j, j = 1, 2, \ldots, J\}$

$\sigma^2_j$ can be generated from the full conditional

$$\left[ \sigma^2_j \mid \text{rest} \right] \propto IG\left(\frac{I^{-1} + \kappa, \left[ \frac{1}{2} (s_j - \mu_j)' \Sigma_j^{-1} (s_j - \mu_j) + \frac{1}{\delta} \right]^{-1}\right)$$

(6) Generate $\kappa$

Random-walk Metropolis-Hasting algorithm is used to generate draws of $\kappa$. Let $\kappa^{(p)}$ be the previous draw for $\kappa_j$, and the next draw $\kappa^{(n)}$ is given by

$$\kappa^{(n)} = \kappa^{(p)} + \Delta \kappa,$$

where $\Delta \kappa$ is a draw from a candidate generating density Normal $(0, 0.0005)$ for the business to business dataset and a candidate generating density Normal $(0, 0.0003)$ for the business to customer dataset. $\kappa$ is any value greater than 3 to grantee the existence of first two moments of inverted gamma distribution. The choice for parameters of this density ensures 30% acceptance rate. Thus, the probability of move $\kappa$ is given by
\[
f(k^{(p)}, k^{(n)}) = \min \left( \frac{\pi(k^{(n)})}{\pi(k^{(p)})}, 1 \right)
\]

, where \( \pi[k] \propto \prod_{j=1}^{J} \frac{1}{\Gamma(\kappa)} \delta^{\kappa} (\sigma_j^2)^{-\kappa-1} \cdot I(\kappa \geq 3) \)

(7) Generate \( \delta \)

Assume the prior distribution \( \delta \sim IG(d_0, D_0) \). \( \delta \) can be generated from

the full conditional

\[
[\delta | \text{rest}] \propto IG \left( \kappa J + d_0, \left[ \frac{1}{\sum_{j=1}^{J} \sigma_j^2} + \frac{1}{D_0} \right]^{-1} \right)
\]

The density of inverted gamma for generating \( \sigma_\delta^2 \), \( \{ \sigma_j^2, j = 1, 2, \cdots, J \} \), \( \kappa \),

and \( \delta \) is

\[
x \sim IG(\alpha, \beta) = \frac{1}{\beta^\alpha \Gamma(\alpha)} (x)^{-(\alpha+1)} \exp \left( -\frac{1}{\beta x} \right)
\]

B.2 Inter-purchase Time Model

Let \( r_{i,j} \) is the inter-purchase time of customer \( j \) between \( i^{th} \) and \( (i+1)^{th} \) order and \( w_{i,j} \) is the purchase amount of customer \( j \) at \( i^{th} \) order. Assuming \( r_{i,j}^* \) is log-normally distributed, that is \( r_{i,j}^* \sim LN \left( \beta_{0,j} + \beta_{i,j} w_j, \sigma_j^2 \right) \), or \( r_{i,j} = \ln \left( r_{i,j}^* \right) \sim N \left( \beta_{0,j} + \beta_{i,j} w_j, \sigma_j^2 \right) \),

\( i = 1, 2, \ldots, n_j \) and \( j = 1, 2, \ldots, J \).
(1) Generate $\beta_j = (\beta_{0,j}, \beta_{1,j})$

Assume the prior distribution $\beta_j = (\beta_{0,j}, \beta_{1,j}) \sim N_2(\bar{\beta}, V_\beta)$. $\beta_j$ can be generated from the full conditional

$$[\beta_j | \text{rest}] \propto N_2\left(\left[\frac{1}{\sigma_j} W_j^T W_j + V_\beta^{-1}\right]^{-1}\left[\frac{1}{\sigma_j} W_j^T n_j + V_\beta^{-1}\bar{\beta}\right], \left[\frac{1}{\sigma_j} W_j^T W_j + V_\beta^{-1}\right]^{-1}\right)$$

where $W_j = \begin{bmatrix} 1 & w_{1,j} \\ \vdots & \vdots \\ 1 & w_{nj,j} \end{bmatrix}$, $r_j = \begin{bmatrix} r_{1,j} \\ \vdots \\ r_{nj,j} \end{bmatrix}$, $\bar{\beta} = (\bar{\beta}_o, \bar{\beta}_i)'$.

(2) Generate $\bar{\beta} = (\bar{\beta}_o, \bar{\beta}_i)'$

Assume the prior distribution $\bar{\beta} = (\bar{\beta}_o, \bar{\beta}_i)' \sim N_2(a_0, S_0)$. $\bar{\beta}$ can be generated from the full conditional

$$[\bar{\beta} | \text{rest}] \propto N_2\left Jehovah(\left[J V_\beta^{-1} + S_0^{-1}\right]^{-1}\left[TV_\beta^{-1}B'B + S_0^{-1}\{a_0\}, [JTV_\beta^{-1} + S_0^{-1}]^{-1}\right)$$

where $B = \begin{bmatrix} \beta_i & \cdots & \beta_j & \cdots & \beta_j \end{bmatrix}$ and $\beta_j = (\beta_{0,j}, \beta_{1,j})'$. 

172
(3) Generate $V_β$

Assume the prior distribution $V_β \sim IW_2(v_0, G_0)$. $V_β$ can be generated from the full conditional

$$
[V_β | \text{rest}] \propto IW_2\left(v + J, G_0 + \sum_{j=1}^{J} (\beta_j - \bar{\beta})(\beta_j - \bar{\beta})'\right)
$$

(4) Generate $σ_j^2$

Assume the prior distribution $σ_j^2 \sim IG(κ, δ)$. $σ_j^2$ can be generated from the full conditional

$$
[σ_j^2 | \text{rest}] \propto IG\left[κ + \frac{n_j}{2}, \left(\frac{1}{2}(r_j - W_jβ_j)'(r_j - W_jβ_j) + δ^{-1}\right)^{-1}\right]
$$

(5) Generate $κ$

The likelihood function is

$$
L(κ|δ, σ_j^2) = \prod_{j=1}^{J} \frac{1}{\Gamma(κ)}δ^κ(σ_j^2)^{-(κ+1)} \exp\left(-\frac{1}{δσ_j^2}\right) \propto \prod_{j=1}^{J} \frac{(δσ_j^2)^{-(κ+1)}}{Γ(κ)} \exp\left(-\frac{1}{δσ_j^2}\right)
$$

Posterior Distribution of $κ$, given $σ_j^2$ and $δ$, is proportional to

$$
\prod_{j=1}^{J} \text{continuous poisson}\left(\frac{1}{δσ_j^2}\right), \text{which is the product of } N \text{ continuous Poisson distributions}.
$$
(6) Generate $\delta$

Assume the prior distribution $\delta \sim IG(d_0, D_0)$. $\delta$ can be generated from the full conditional

$$[\delta \mid \text{rest}] \sim IG\left(\kappa J + d_0, \left(\sum_{j=1}^{j} \sigma_j^{-2} + D_0^{-1}\right)^{-1}\right)$$
APPENDIX C

MCMC ESTIMATION FOR ESSAY 3
The proposed model with multiplicative and additive consumer heterogeneity and two alternative models follow the same solution procedure. The solution procedure is summarized as follows:

**Step 1:** Search the optimal quantity of each brand for all feasible brand-pack combinations

**Step 2:** Generate latent utilities for the items with optimal quantity according to the multinomial probit model

**Step 3:** Estimate parameters using the proposed error augmentation and check-and-see algorithm. Check-and-see algorithm is performed by checking the following conditions

\[ I \left( \ln u \left( x_{th}^* \right) \right) = 1 \text{ if } \ln u(x_{kbt} = x_{th}^*) > \max \left\{ \ln u \left( x_{ti}^* \right), \forall b' \neq b \right\} \text{ and } x_{kbt} \neq 0 \]

\[ I \left( \ln u \left( x_{it}^* \right) \right) = 1 \text{ if } \max \left\{ \ln u \left( x_{i,t}^* = 0 \right) \right\} > \max \left\{ \ln u \left( x_{i,t}^* \neq 0 \right) \right\} \text{ and } \{i' \cup i^*\} = \{i\} \]

The procedure of generating latent utilities illustrated in Section 3.2 is applied to all four models. In this appendix, only the conditional distributions of parameters except generating latent utilities are provided.

**C.1 Baseline Model**

Pretest and posttest data are fit respectively by the model specified in Equations (5.51) and (5.52) with the same prior assumptions. Let \( \ln U_{th} \) denote the vector of log latent utilities of all items available to respondent \( h \) at the shopping trip \( t \), and \( \ln u \left( x_{th}^* \right) \) denote the vector of log latent utilities of all items included in the optimal set of respondent \( h \) at the shopping trip \( t \).
(1) Generate

\[ \ln u(x_{ihb}), \forall i = 1, 2, \ldots, K; h = 1, 2, \ldots, H; t = 1, 2, \ldots, T; b = 2, 3, \ldots, B; t = 1, 2, \ldots, T_1 \]

for pretest data. Generate \( \ln u(x_{ihb}), \forall i = 1, 2, \ldots, K; h = 1, 2, \ldots, H; t = 1, 2, \ldots, T; b = 1, 2, \ldots, B; t = 1, 2, \ldots, T_2 \) for pretest data. Then retain the error realizations.

(2) Generate \( \beta_h, \forall h = 1, 2, \ldots, H \)

\[
\begin{align*}
[\beta_h | \text{rest}] & \propto \prod_{t=1}^{T} I(\ln U_{ih}) \cdot [\ln U_{ih} | \beta_h] \cdot [\beta_h | \bar{\beta}, V_{\beta}] \\
& \propto \prod_{t=1}^{T} I\left(\ln u\left(x_{ih}^*\right)\right) \cdot \left[\ln u\left(x_{ih}^*\right) | \beta_h\right] \cdot [\beta_h | \bar{\beta}, V_{\beta}] \\
& \propto \exp\left\{-\frac{1}{2} \sum_{t=1}^{T} \left[\ln u\left(x_{ih}^*\right) - z_{ih}^* \beta_h\right]^\top \Sigma^{-1} \left[\ln u\left(x_{ih}^*\right) - z_{ih}^* \beta_h\right]\right\} \\
& \cdot \exp\left\{-\frac{1}{2} \left[\beta_h - \bar{\beta}\right]^\top V_{\beta}^{-1} \left[\beta_h - \bar{\beta}\right]\right\} \cdot \prod_{t=1}^{T} I\left(\ln u\left(x_{ih}^*\right)\right)
\end{align*}
\]

Random-walk Metropolis-Hasting algorithm is used to generate the draw of \( \beta_h \). Let \( \beta_h^{(o)} \) be the previous draw for \( \beta_h \), and the next draw \( \beta_h^{(n)} \) is given by

\[ \beta_h^{(n)} = \beta_h^{(o)} + \Delta \beta, \]

where \( \Delta \beta \) is a draw from a candidate generating density Normal \((0, \sigma_{\beta h}^2)\). Let \( \ln u\left(x_{ih}^{*(o)}\right) \) be the previous draw for log utility \( \ln u\left(x_{ih}^*\right) \), and the next draw \( \ln u\left(x_{ih}^{*(n)}\right) \) is given by

\[ \ln u(x_{ih}^{*(n)}) = z_{ih}^* \beta_h^{(n)} + e_{ih}^*. \]
The parameter $\sigma_{\beta_h}^2$ of this density is adjusted in each MCMC iteration to ensure 30% acceptance rate. Thus, the probability of move $\beta_h$ is given by

$$\kappa(\beta^{(o)}_h, \beta^{(n)}_h) = \min\left(\frac{\pi(\beta^{(n)}_h)}{\pi(\beta^{(o)}_h)}, 1\right)$$

and

$$\frac{\pi(\beta^{(n)}_h)}{\pi(\beta^{(o)}_h)} = \exp\left\{-\frac{1}{2} \left[\beta^{(n)}_h - \bar{\beta}\right]' V_\beta^{-1} \left[\beta^{(o)}_h - \bar{\beta}\right]\right\} \cdot \prod_{t=1}^{T} I\left(\ln u\left(x^{(n)}_{ih}\right)\right)$$

$$\propto \prod_{t=1}^{T} I\left(\ln u\left(x^{(o)}_{ih}\right)\right)$$

(3) Generate $T_h$, $\forall h = 1, 2, \ldots, H$

$$[T_h \mid \text{rest}] \propto \prod_{t=1}^{T} I\left(\ln U_{ih}\right) \cdot [\ln U_{ih} \mid T_h] \cdot [T_h \mid T, \sigma_r^2]$$

$$\propto \prod_{t=1}^{T} I\left(\ln u\left(x^{*}_{ih}\right)\right) \cdot [\ln u\left(x^{*}_{ih}\right) \mid T_h] \cdot [T_h \mid T, \sigma_r^2]$$

$$\propto \exp\left\{-\frac{1}{2} \sum_{t=1}^{T} \left[\ln u\left(x^{*}_{ih}\right) - z^{*}_{ih} \beta_h\right]' \Sigma^{-1} \left[\ln u\left(x^{*}_{ih}\right) - z^{*}_{ih} \beta_h\right]\right\} \cdot \exp\left\{-\frac{1}{2\sigma_r^2} \left[T_h - \bar{T}\right]^2\right\} \cdot \prod_{t=1}^{T} I\left(\ln u\left(x^{*}_{ih}\right)\right)$$

$$\propto \exp\left\{-\frac{1}{2} \sum_{t=1}^{T} e^{*}_{ih}' \Sigma^{-1} e^{*}_{ih}\right\} \cdot \exp\left\{-\frac{1}{2\sigma_r^2} \left[T_h - \bar{T}\right]^2\right\} \cdot \prod_{t=1}^{T} I\left(\ln u\left(x^{*}_{ih}\right)\right)$$

Random-walk Metropolis-Hasting algorithm is used to generate the draw of $T_h$. Let $T_h^{(o)}$ be the previous draw for $T_h$, and the next draw $T_h^{(n)}$ is given by

$$T_h^{(n)} = T_h^{(o)} + \Delta T$$
where $\Delta T$ is a draw from a candidate generating density Normal $(0, 1)$. The choice for parameters of this density ensures 30% acceptance rate. Let $\ln u^{(o)}(x_{th}^*)$ be the previous draw for log utility $\ln u(x_{th}^*)$, and the next draw $\ln u^{(n)}(x_{th}^*)$ is given by

$$
\ln u^{(n)}(x_{th}^*) = z_{ht}^{(n)} r + e_{th}^*.
$$

Thus, the probability of move $T_h$ is given by

$$
\kappa(T_h^{(o)}, T_h^{(n)}) = \min \left\{ \frac{\pi(T_h^{(n)})}{\pi(T_h^{(o)})}, 1\right\}
$$

and

$$
\frac{\pi(T_h^{(n)})}{\pi(T_h^{(o)})} = \exp \left\{-\frac{1}{2\sigma_T^2} \left[ T_h^{(n)} - T_h^{(o)} \right]^2\right\} \cdot \prod_{i=1}^T \left[ I \left( \ln u^{(n)}(x_{hi}^*) \right) \right]
$$

(4) Generate $\bar{\beta}$

$$
\begin{bmatrix} \bar{\beta} | \text{rest} \end{bmatrix} \propto \prod_{h=1}^H \begin{bmatrix} \beta_h | \bar{\beta}, V_{\beta} \end{bmatrix} \cdot \begin{bmatrix} \bar{\beta} | \mu_0, V_0 \end{bmatrix}
$$

$$
\sim N \left( \begin{bmatrix} H V_{\beta}^{-1} + V_0^{-1} \end{bmatrix}^{-1} \left[ V_{\beta}^{-1} \sum_{h=1}^H \beta_h + V_0^{-1} \mu_0 \right], \begin{bmatrix} H V_{\beta}^{-1} + V_0^{-1} \end{bmatrix}^{-1} \right)
$$

(5) Generate $V_{\beta}$

$$
\begin{bmatrix} V_{\beta} | \text{rest} \end{bmatrix} \propto \prod_{h=1}^H \begin{bmatrix} \beta_h | \bar{\beta}, V_{\beta} \end{bmatrix} \cdot \begin{bmatrix} V_{\beta} | \nu, G \end{bmatrix}
$$

$$
\sim IW \left( G + \nu + nb \text{ var.} \sum_{h=1}^H (\beta_h - \bar{\beta})(\beta_h - \bar{\beta})' + G \right)
$$
(6) Generate $T$

$$
\begin{bmatrix} T | \text{rest} \end{bmatrix} \propto \prod_{h=1}^{H} \begin{bmatrix} T_h | T, \sigma_T^2 \end{bmatrix} \cdot \begin{bmatrix} T | a, b^2 \end{bmatrix} \\
\sim \mathcal{N} \left( \frac{H}{\sigma_T^2 + \frac{1}{b^2}} \begin{bmatrix} \sum_{h=1}^{H} T_h \end{bmatrix}^{-1} \begin{bmatrix} H \sigma_T^2 + \frac{1}{b^2} \end{bmatrix} \right)
$$

(7) Generate $\sigma_T^2$

$$
\begin{bmatrix} \sigma_T^2 | \text{rest} \end{bmatrix} \propto \prod_{h=1}^{H} \begin{bmatrix} T_h | T, \sigma_T^2 \end{bmatrix} \cdot \begin{bmatrix} \sigma_T^2 | \nu \sigma_0, s_0^2 \end{bmatrix} \\
\sim \frac{\nu \sigma_1^2}{\chi^2_{\nu_1}}, \text{where } \nu_1 = H + \nu \sigma_0 \text{ and } \nu \sigma_1^2 = \sum_{h=1}^{H} (T_h - \bar{T})^2 + \nu \sigma_0^2
$$

C.2 Two Random-effect Model

Pretest and posttest data are fit by the model specified in Equations (5.18), (5.20), (5.53), and (5.54). Let $\ln U^{(1)}_{th}$ denote the vector of pretest latent utilities of all items available to respondent $h$ at the pretest shopping trip $t$, and $\ln u(\mathbf{x}^{(1)*}_{th})$ denote the vector of pretest latent utilities of all items included in the optimal set of respondent $h$ at the pretest shopping trip $t$. Let $\ln U^{(2)}_{th}$ denote the vector of posttest latent utilities of all items available to respondent $h$ at the posttest shopping trip $t$, and $\ln u(\mathbf{x}^{(2)*}_{th})$ denote the vector of posttest latent utilities of all items included in the optimal set of respondent $h$ at the pretest shopping trip $t$. 

180
(1) Generate

\[
\ln u \left( x_{iabt}^{(1)} \right), \forall i = 1, 2, \ldots, K; b = 2, 3, \ldots, B; h = 1, 2, \ldots, H; t = 1, 2, \ldots, T_i,
\]

\[
\ln u \left( x_{iabt}^{(2)} \right), \forall i = 1, 2, \ldots, K; b = 1, 2, \ldots, B; h = 1, 2, \ldots, H; t = 1, 2, \ldots, T_2. \text{ Then retain the error realizations.}
\]

(2) Generate \( \beta_h^{(1)}, \forall h = 1, 2, \ldots, H \)

Since \( \beta_h^{(1)} = \left( \beta_{2h}, \beta_{th} \right)^{(1)^T} = \left( \beta_{02h}, \beta_{03h}, \ldots, \beta_{0Bh}, \beta_{z_1h}, \ldots, \beta_{z_nvarh}, \beta_{xh}, \beta_{\theta h} \right)^{(1)^T} \),

\[
\beta_h^{(1)^*} = \left( \beta_{2h}, \beta_{\theta h}^{*} \right)^{(1)^T} = \left( \beta_{02h}, \beta_{03h}, \ldots, \beta_{0Bh}, \beta_{z_1h}^*, \ldots, \beta_{z_nvarh}, \beta_{xh}, \beta_{\theta h}^* \right)^{(1)^T}, \text{ and}
\]

\[
\beta_{\theta h}^{(1)} = \exp \left( \beta_{\theta h}^{(1)^*} \right), \text{ the conditional distribution of } \beta_h^{(1)} \text{ is}
\]

\[
\left[ \beta_h^{(1)} \mid \text{rest} \right] \propto \prod_{i=1}^{T_i} \left[ I \left( \ln U_{ih}^{(1)} \right) \cdot \ln U_{ih}^{(1)} \mid \beta_h^{(1)} \right] \cdot \left[ \beta_h^{(1)^*} \mid \beta_h^{(1)}, V_i^{(1)} \right]
\]

\[
\propto \prod_{i=1}^{T_i} \left[ I \left( \ln u \left( x_{ih}^{(1)^*} \right) \right) \cdot \ln u \left( x_{ih}^{(1)^*} \right) \mid \beta_h^{(1)} \right] \cdot \left[ \beta_h^{(1)^*} \mid \beta_h^{(1)}, V_i^{(1)} \right]
\]

\[
\propto \exp \left\{ -\frac{1}{2} \sum_{i=1}^{T_i} \left[ \ln u \left( x_{ih}^{(1)^*} \right) - \left( x_{ih}^{(1)^*} - \beta_{\theta h}^{(1)} \beta_{\theta h}^{(1)^*} \right)^{\Sigma^{(1)^{-1}}} \ln u \left( x_{ih}^{(1)^*} \right) - \beta_{\theta h}^{(1)^*} \right] \right\}.
\]

\[
\exp \left\{ -\frac{1}{2} \left[ \beta_h^{(1)^*} - \bar{\theta}^{(1)} \right]^T \left( V_\beta^{(1)} \right)^{-1} \left[ \beta_h^{(1)^*} - \bar{\theta}^{(1)} \right] \right\} \prod_{i=1}^{T_i} I \left( \ln u \left( x_{ih}^{(1)^*} \right) \right)
\]

\[
\propto \exp \left\{ -\frac{1}{2} \sum_{i=1}^{T_i} e_{ih}^{(1)^*} \Sigma^{(1)^{-1}} e_{ih}^{(1)^*} \right\}.
\]

\[
\exp \left\{ -\frac{1}{2} \left[ \beta_h^{(1)^*} - \bar{\theta}^{(1)} \right]^T \left( V_\beta^{(1)} \right)^{-1} \left[ \beta_h^{(1)^*} - \bar{\theta}^{(1)} \right] \right\} \prod_{i=1}^{T_i} I \left( \ln u \left( x_{ih}^{(1)^*} \right) \right)
\]

Random-walk Metropolis-Hasting algorithm is used to generate the draw of \( \beta_h^{(1)} \).

Let \( \beta_h^{(1)^{(o)}} \) be the previous draw for \( \beta_h^{(1)^*} \), and the next draw \( \beta_h^{(1)^{(n)}} \) is given by

\[
\beta_h^{(1)^{(n)}} = \beta_h^{(1)^{(o)}} + \Delta \beta,
\]

181
where $\Delta \beta$ is a draw from a candidate generating density Normal $(0, \sigma_{\beta}^2 h)$. Let

$$\ln u^{(o)}(x_{th})$$

be the previous draw for log utility $\ln u(x_{th}^{(s)})$, and the next draw $\ln u^{(a)}(x_{th})$ is given by

$$\ln u^{(a)}(x_{th}^{(s)}) = z^*_h \beta_h^{(1)(a)} + e^*_h.$$

The parameter $\sigma_{\beta_h}^2$ of this density is adjusted in each MCMC iteration to ensure 30% acceptance rate. Thus, the probability of move $\beta_h^{(1)r}$ is given by

$$\kappa(\beta_h^{(1)r(o)}, \beta_h^{(1)r(n)}) = \min \left( \frac{\pi(\beta_h^{(1)r(o)})}{\pi(\beta_h^{(1)r(n)})}, 1 \right)$$

and

$$\frac{\pi(\beta_h^{(1)r(o)})}{\pi(\beta_h^{(1)r(n)})} = \frac{\exp \left\{ -\frac{1}{2} \left[ \beta_h^{(1)r(o)} - \bar{\beta}_h^{(1)} \right]' \left( V^{(1)}_h \right)^{-1} \left[ \beta_h^{(1)r(o)} - \bar{\beta}_h^{(1)} \right] \right\} \prod_{t=1}^{T} I \left( \ln u^{(a)}(x_{th}^{(s)}) \right)}{\exp \left\{ -\frac{1}{2} \left[ \beta_h^{(1)r(n)} - \bar{\beta}_h^{(1)} \right]' \left( V^{(1)}_h \right)^{-1} \left[ \beta_h^{(1)r(n)} - \bar{\beta}_h^{(1)} \right] \right\} \prod_{t=1}^{T} I \left( \ln u^{(o)}(x_{th}^{(s)}) \right)}.$$
(3) Generate $\beta_h^{(2)}$, $\forall h = 1, 2, ..., H$

\[
[\beta_h^{(2)} \mid \text{rest}] \propto \prod_{i=1}^{T_h} I\left(\ln U_{ih}\right) \cdot \left[\ln U_{ih} \mid \beta_h^{(2)}\right] \cdot \left[\beta_h^{(2)*} \mid \beta_h^{(2)}, \nu_{\beta}^{(2)}\right] \\
\propto \prod_{i=1}^{T_h} I\left(\ln u\left(x_{ih}^{(2)*}\right)\right) \cdot \left[\ln u\left(x_{ih}^{(2)*}\right) \mid \beta_h^{(2)}\right] \cdot \left[\beta_h^{(2)*} \mid \beta_h^{(2)}, \nu_{\beta}^{(2)}\right] \\
\propto \exp\left\{-\frac{1}{2} \sum_{t=1}^{T_h} \left[\ln u\left(x_{ih}^{(2)*}\right) - \nu_{\beta h}^{(2)*} - \beta_h^{(2)}\right]^{\nu_{\beta h}^{(2)*}}\right\} \\
\exp\left\{-\frac{1}{2} \left[\beta_h^{(2)*} - \beta_h^{(2)}\right]^{\nu_{\beta}^{(2)*}}\right\} \\
\exp\left\{-\frac{1}{2} \left[\beta_h^{(2)*} - \beta_h^{(2)}\right]^{\nu_{\beta}^{(2)*}}\right\} \\
\prod_{i=1}^{T_h} I\left(\ln u\left(x_{ih}^{(2)*}\right)\right)
\]

Random-walk Metroplis-Hasting algorithm is used to generate the draw of $\beta_h^{(2)}$.

Let $\beta_h^{(2)*^{(o)}}$ be the previous draw for $\beta_h^{(2)*}$, and the next draw $\beta_h^{(2)*^{(n)}}$ is given by

$$
\beta_h^{(2)*^{(n)}} = \beta_h^{(2)*^{(o)}} + \Delta \beta,
$$

where $\Delta \beta$ is a draw from a candidate generating density Normal $(0, \sigma_{\beta h}^{(2)*})$. Let

$\ln u^{(o)}\left(x_{ih}^{(2)*}\right)$ be the previous draw for log utility $\ln u\left(x_{ih}^{(2)*}\right)$, and the next draw $\ln u^{(n)}\left(x_{ih}^{(2)*}\right)$ is given by

$$
\ln u^{(n)}\left(x_{ih}^{(2)*}\right) = z_h^* \beta_h^{(2)*^{(n)}} + e_{ih}^*.
$$

The parameter $\sigma_{\beta h}^{(2)*}$ of this density is adjusted in each MCMC iteration to ensure 30% acceptance rate. Thus, the probability of move $\beta_h^{(2)*}$ is given by

$$
\kappa(\beta_h^{(2)*^{(o)}}, \beta_h^{(2)*^{(n)}}) = \min\left\{\frac{\pi\left(\beta_h^{(2)*^{(n)}}\right)}{\pi\left(\beta_h^{(2)*^{(o)}}\right)}, 1\right\}
$$

183
and

\[
\frac{\pi(\beta_h^{(2)(o)})}{\pi(\beta_h^{(2)(o)(o)})} = \exp \left\{ -\frac{1}{2} \left[ \beta_h^{(2)(o)(o)} - \bar{\beta}^{(2)} \right] (Y_\beta^{(2)})^{-1} \left[ \beta_h^{(2)(o)(o)} - \bar{\beta}^{(2)} \right] \right\} \cdot \prod_{t=1}^{T} I\left( u^{(o)}(x_{th}^{(1)}) \right)
\]

(4) Generate \( T_h, \forall h = 1, 2, \ldots, H \)

\[
[T_h \mid \text{rest}] \propto \prod_{t=1}^{T} I\left( \ln U_{th}^{(1)} \right) \cdot \prod_{t=1}^{T} I\left( \ln U_{th}^{(2)} \right) \cdot \prod_{t=1}^{T} I\left( U_{th}^{(2)} | \beta_h^{(2)} \right) \cdot \prod_{t=1}^{T} I\left( U_{th}^{(2)} | T_h, \bar{T}, \sigma_T^2 \right)
\]

\[
\propto \prod_{t=1}^{T} I\left( u(x_{th}^{(1)}) \right) \cdot \prod_{t=1}^{T} I\left( u(x_{th}^{(2)}) \right) \cdot \prod_{t=1}^{T} I\left( u(x_{th}^{(2)}) \right) \cdot \prod_{t=1}^{T} I\left( x_{th}^{(2)} \right) \cdot [T_h | \bar{T}, \sigma_T^2]
\]

\[
\propto \exp \left\{ -\frac{1}{2} \sum_{t=1}^{T} \left[ \ln u(x_{th}^{(1)}) - z_{th}^{(1)*} \beta_h^{(1)} \right] \Sigma^{(1)^{-1}} \left[ \ln u(x_{th}^{(1)}) - z_{th}^{(1)*} \beta_h^{(1)} \right] \right\}
\]

\[
\exp \left\{ -\frac{1}{2} \sum_{t=1}^{T} \left[ \ln u(x_{th}^{(2)}) - z_{th}^{(2)*} \beta_h^{(2)} \right] \Sigma^{(2)^{-1}} \left[ \ln u(x_{th}^{(2)}) - z_{th}^{(2)*} \beta_h^{(2)} \right] \right\}
\]

\[
\exp \left\{ -\frac{1}{2} \sum_{t=1}^{T} \left[ T_h - \bar{T} \right] \cdot \prod_{t=1}^{T} I\left( \ln u(x_{th}^{(1)}) \right) \cdot \prod_{t=1}^{T} I\left( \ln u(x_{th}^{(2)}) \right) \right\}
\]

\[
\alpha \exp \left\{ -\frac{1}{2} \sum_{t=1}^{T} \left[ T_h - \bar{T} \right] \cdot \prod_{t=1}^{T} I\left( \ln u(x_{th}^{(1)}) \right) \cdot \prod_{t=1}^{T} I\left( \ln u(x_{th}^{(2)}) \right) \right\}
\]

Random-walk Metropolis-Hasting algorithm is used to generate the draw of \( T_h \). Let \( T_h^{(o)} \) be the previous draw for \( T_h \), and the next draw \( T_h^{(n)} \) is given by

\[
T_h^{(n)} = T_h^{(o)} + \Delta T,
\]

where \( \Delta T \) is a draw from a candidate generating density Normal (0, 1). Let

\[
\ln u^{(o)}(x_{th}^{(1)}) \text{ be the previous draw for log utility } \ln u(x_{th}^{(1)}) \text{, and the next draw }
\]

\[
\ln u^{(n)}(x_{th}^{(1)}) \text{ is given by }
\]

184
\[ \ln u^{(n)}(x_{ht}^{(1)*}) = z_{ht}^{(n)}* \beta_{h}^{(1)} + e_{ht}^*. \]

Let \( \ln u^{(o)}(x_{th}^{2*y}) \) be the previous draw for log utility \( \ln u(x_{th}^{2*y}) \), and the next draw \( \ln u^{(n)}(x_{th}^{2*y}) \) is given by

\[ \ln u^{(n)}(x_{ht}^{2*y}) = z_{ht}^{(n)}* \beta^{(2)*} + e_{ht}^*. \]

The choice for parameters of this density ensures 30% acceptance rate. Thus, the probability of move \( T_h \) is given by

\[ \kappa(T_h^{(o)}, T_h^{(n)}) = \min \left\{ \frac{\pi\left(T_h^{(n)}\right)}{\pi\left(T_h^{(o)}\right)}, 1 \right\} \]

and

\[ \frac{\pi\left(T_h^{(o)}\right)}{\pi\left(T_h^{(n)}\right)} = \exp\left\{ \frac{1}{2\sigma_T^2} [T_h^{(o)} - \overline{T}]^2 \right\} \cdot \prod_{t=1}^{T_o} I\left( \ln u^{(o)}(x_{th}^{(1)*}) \right) \cdot \prod_{t=1}^{T_n} I\left( \ln u^{(n)}(x_{th}^{2*y}) \right) \]

(5) Generate \( \Gamma = (\gamma_{01}^{(n)}, \gamma_{02}^{(n)}, \ldots, \gamma_{r}^{(n)}) \)

\[ \begin{bmatrix} \text{vec}(\Gamma) | \text{rest} \end{bmatrix} \propto \prod_{h=1}^{H} \left[ \beta_h^{(2)*} \mid \overline{\beta}_h^{(2)}, V_\beta \right] \begin{bmatrix} \text{vec}(\Gamma) \mid \overline{\gamma}, V_\gamma \end{bmatrix} \]

\[ \propto \exp\left\{ -\frac{1}{2} \sum_{h=1}^{H} \left( \beta_h^{(2)*} - \overline{\beta}_h^{(2)} \right)' V_\beta^{-1} \left( \beta_h^{(2)*} - \overline{\beta}_h^{(2)} \right) \right\} \]

\[ \exp\left\{ -\frac{1}{2} \left( \text{vec}(\Gamma - \overline{\gamma}) \right)' (I \otimes V_\gamma)^{-1} \left( \text{vec}(\Gamma - \overline{\gamma}) \right) \right\} \]

Random-walk Metroplis-Hasting algorithm is used to generate the draw of \( \Gamma \). Let \( \Gamma^{(o)} \) be the previous draw for \( \Gamma_h \), and the next draw \( \Gamma_h^{(n)} \) is given by

\[ \Gamma^{(n)} = \Gamma^{(o)} + \Delta \gamma \cdot \sigma_{\beta_h^{(2)}} \]
where $\Delta \gamma$ is a draw from a candidate generating density $\text{Normal}(0, \sigma_{\gamma h}^2)$. Let $\beta^{(2)(o)}$ be the previous draws for $\beta^{(2)}$, and the next draw $\beta^{(2)(n)}$ is given by

$$
\beta^{(2)(n)} = \begin{bmatrix}
\beta_{01}^{(1)} + \gamma_{01}^{(n)} m_{01,h} \\
\beta_{02}^{(1)} + \gamma_{02}^{(n)} m_{02,h} \\
\beta_{04}^{(1)} + \gamma_{04}^{(n)} m_{04,h} \\
\beta_{05}^{(1)} + \gamma_{05}^{(n)} m_{05,h} \\
\beta_{06}^{(1)} + \gamma_{06}^{(n)} m_{05,h} \\
\beta_{e_x}^{(1)} + \gamma_{e_x}^{(n)} M_{h} \\
\beta_{e_T}^{(1)} + \gamma_{e_T}^{(n)} M_{h} \\
\end{bmatrix}
$$

The choice for parameters $\sigma_{\gamma h}^2$ of this density is adjusted in each MCMC iteration to ensure 30% acceptance rate. Let $\ln u^{(n)}(x_{th}^{(2)*})$ be the previous draw for log utility $\ln u(x_{th}^{(2)*})$, and the next draw $\ln u^{(n)}(x_{th}^{(2)*})$ is given by

$$
\ln u^{(n)}(x_{th}^{(2)*}) = z_{th}^* \beta_{h}^{(2)(n)} + e_{h}^*.
$$

Thus, the probability of move $\Gamma$ is given by

$$
\kappa(\Gamma^{(o)}, \Gamma^{(n)}) = \min \left( \frac{\pi(\Gamma^{(n)})}{\pi(\Gamma^{(o)})}, 1 \right)
$$

and

186
\[
\pi \left( \Gamma^{(n)} \right) = \frac{\exp \left\{ -\frac{1}{2} \sum_{h=1}^{H} \left( \beta_{h}^{(2)*} - \overline{\beta}_{h}^{(2)(n)} \right)' V_{\gamma}^{(2)*}^{-1} \left( \beta_{h}^{(2)*} - \overline{\beta}_{h}^{(2)(n)} \right) \right\}}{\exp \left\{ -\frac{1}{2} \left( \text{vec} \left( \Gamma^{(o)} - \overline{\gamma} \right) \right)' \left( I \otimes V_{\gamma} \right)^{-1} \left( \text{vec} \left( \Gamma^{(o)} - \overline{\gamma} \right) \right) \right\}}.
\]

(6) Generate \( \overline{\beta}^{(1)} \)

\[
\left[ \overline{\beta}^{(1)} \mid \text{rest} \right] \propto \prod_{h=1}^{H} \left[ \beta_{h}^{(1)*} \mid \overline{\beta}^{(1)}, V_{\beta}^{(1)} \right] \cdot \prod_{h=1}^{H} \left[ \beta_{h}^{(2)*} \mid \overline{\beta}_{h}^{(2)}, V_{\beta}^{(2)} \right] \left[ \overline{\beta}^{(1)} \mid \mu_{0}, V_{0} \right]
\]

\[
\propto \exp \left\{ -\frac{1}{2} \sum_{h=1}^{H} \left( \beta_{h}^{(1)*} - \overline{\beta}^{(1)} \right)' V_{\beta}^{(1)*}^{-1} \left( \beta_{h}^{(1)*} - \overline{\beta}^{(1)} \right) \right\}.
\]

\[
\exp \left\{ -\frac{1}{2} \sum_{h=1}^{H} \beta_{h}^{(2)*}' V_{\beta}^{(2)*}^{-1} \left( \beta_{h}^{(2)*} - \overline{\beta}_{h}^{(2)} \right) \right\}.
\]

\[
\exp \left\{ -\frac{1}{2} \left( \overline{\beta}^{(1)} - \mu_{0} \right)' V_{0}^{-1} \left( \overline{\beta}^{(1)} - \mu_{0} \right) \right\}
\]

Random-walk Metropolis-Hasting algorithm is used to generate the draw of \( \overline{\beta}^{(1)} \).

Let \( \overline{\beta}^{(1)(o)} \) be the previous draw for \( \overline{\beta}^{(1)} \), and the next draw \( \overline{\beta}^{(1)(n)} \) is given by

\[
\overline{\beta}^{(1)(n)} = \overline{\beta}^{(1)(o)} + \Delta \overline{\beta},
\]

where \( \Delta \overline{\beta} \) is a draw from a candidate generating density Normal \( (0, \sigma_{\beta}^{2} h) \). Let

\( \overline{\beta}^{(2)(o)} \) be the previous draws for \( \overline{\beta}^{(2)} \), and the next draw \( \overline{\beta}^{(2)(n)} \) is given by
The parameter $\sigma_{\beta_h}^2$ of this density is adjusted in each MCMC iteration to ensure 30\% acceptance rate. Thus, the probability of move $\tilde{\beta}^{(1)}$ is given by

$$\kappa(\tilde{\beta}^{(1)(o)}, \tilde{\beta}^{(1)(e)}) = \min \left\{ \frac{\pi(\tilde{\beta}^{(1)(e)})}{\pi(\tilde{\beta}^{(1)(o)})}, 1 \right\}$$

and

$$\frac{\pi\left(\tilde{\beta}^{(1)(o)}\right)}{\pi\left(\tilde{\beta}^{(1)(e)}\right)} = \exp\left\{ -\frac{1}{2} \sum_{k=1}^{H} \left( \beta_h^{(1)*} - \tilde{\beta}_h^{(1)(e)} \right)' V_\beta^{(1)-1} \left( \beta_h^{(1)*} - \tilde{\beta}_h^{(1)(e)} \right) \right\}.$$

$$\exp\left\{ -\frac{1}{2} \sum_{h=1}^{H} \left( \beta_h^{(2)*} - \tilde{\beta}_h^{(2)(o)} \right)' V_\beta^{(2)-1} \left( \beta_h^{(2)*} - \tilde{\beta}_h^{(2)(o)} \right) \right\}.$$
(7) Generate $V^{(1)}_\beta$

$$
\left[ V^{(1)}_\beta | \text{rest} \right] \propto \prod_{h=1}^H \left[ \beta_h^{(1)*} | \bar{\beta}^{(1)}_{\beta}, V^{(1)}_\beta \right] \cdot \left[ V^{(1)}_\beta | \nu_1, G_1 \right]
\sim I W \left( H + \nu_1, \sum_{h=1}^H \left( \beta_h^{(1)*} - \bar{\beta}^{(1)} \right) \left( \beta_h^{(1)*} - \bar{\beta}^{(1)} \right)' + G_1 \right)
$$

(8) Generate $V^{(2)}_\beta$

$$
\left[ V^{(2)}_\beta | \text{rest} \right] \propto \prod_{h=1}^H \left[ \beta_h^{(2)*} | \bar{\beta}^{(2)}_\beta, V^{(2)}_\beta \right] \cdot \left[ V^{(2)}_\beta | \nu_2, G_2 \right]
\sim I W \left( H + \nu_2, \sum_{h=1}^H \left( \beta_h^{(2)*} - \bar{\beta}^{(2)} \right) \left( \beta_h^{(2)*} - \bar{\beta}^{(2)} \right)' + G_2 \right)
$$

(9) Generate $\bar{\gamma}$

$$
\left[ \bar{\gamma} | \text{rest} \right] \propto \prod_{i=1}^{nb \text{var}^2} \left[ \gamma_i | \bar{\gamma}, V_{\gamma} \right] \cdot \left[ \bar{\gamma} | \mu_{\gamma}, \Sigma_{\gamma} \right]
\propto \exp \left\{ -\frac{1}{2} \sum_{i=1}^{nb \text{var}^2} \left( \gamma_i - \bar{\gamma} \right)' V_{\gamma}^{-1} \left( \gamma_i - \bar{\gamma} \right) \right\},
\exp \left\{ -\frac{1}{2} \exp \left( \bar{\gamma} - \mu_{\gamma} \right)' \Sigma^{-1}_{\gamma} \left( \bar{\gamma} - \mu_{\gamma} \right) \right\}
\sim N \left( \left[ nb \text{ var} \cdot V_{\gamma}^{-1} + \Sigma^{-1}_{\gamma} \right]^{-1} \left[ V_{\gamma}^{-1} \left( \sum_{i=1}^{nb \text{ var}^2} \gamma_i \right) + \Sigma^{-1}_{\gamma} \mu_{\gamma} \right], \left[ nb \text{ var} \cdot V_{\gamma}^{-1} + \Sigma^{-1}_{\gamma} \right]^{-1} \right)
$$
(10) Generate $V_\gamma$
\[
[V_\gamma \mid \text{rest}] \propto \prod_{i=1}^{\text{nb var}} \left[ V_\gamma \right] \prod_{i=1}^{\text{var}} \left[ \gamma_i \mid f_{0}, G_0 \right] \prod_{i=1}^{\text{var}} \left[ V_\gamma \mid f_{0}, G_0 \right]
\]
\[
\propto \frac{1}{V_\gamma^{-1}} \exp \left\{ -\frac{1}{2} \text{tr} \left( V_\gamma^{-1} \sum_{i=1}^{\text{nb var}} \left( \gamma_i - \bar{\gamma} \right) \left( \gamma_i - \bar{\gamma} \right)' \right) \right\}
\]
\[
\propto \frac{1}{V_\gamma^{-1}} \exp \left\{ -\frac{1}{2} \text{tr} \left( V_\gamma^{-1} G_0 \right) \right\}
\]
\[
\sim N \left( \begin{bmatrix} nb \text{ var} 2 \cdot V_\gamma^{-1} + \Sigma_\gamma^{-1} \end{bmatrix}^{-1} \begin{bmatrix} V_\gamma^{-1} \sum_{i=1}^{\text{var}} \gamma_i + \Sigma_\gamma^{-1} \mu_\gamma \end{bmatrix} \right) \]

(11) Generate $\bar{T}$
\[
[\bar{T} \mid \text{rest}] \propto \prod_{h=1}^{H} \left[ T_h \mid \bar{T}, \sigma_\gamma^2 \right] \cdot \left[ \bar{T} \mid a, b^2 \right]
\]
\[
\sim N \left( \begin{bmatrix} H \sigma_\gamma^2 + a \sigma_\gamma^2 + 1 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{h=1}^{H} T_h \sigma_\gamma^2 + b \sigma_\gamma^2 + 1 \end{bmatrix} \right)
\]

(12) Generate $\sigma_\gamma^2$
\[
[\sigma_\gamma^2 \mid \text{rest}] \propto \prod_{h=1}^{H} \left[ T_h \mid \bar{T}, \sigma_\gamma^2 \right] \cdot \left[ \sigma_\gamma^2 \mid \nu_0, s_0^2 \right]
\]
\[
\sim \nu_0 s_0^2 \chi^2_{\nu_1}, \text{where } \nu_1 = H + \nu_0 \text{ and } \nu_1 s_0^2 = \sum_{h=1}^{H} \left( T_h - \bar{T} \right)^2 + \nu_0 s_0^2
\]
C.3 Multiplicative Model

Pretest and posttest data are fit by the model specified in Equations (5.18), (5.20), (5.34)-(5.36). Let $\ln U_{ih}^{(1)}$ denote the vector of pretest latent utilities of all items available to respondent $h$ at the pretest shopping trip $t$, and $\ln u\left(x_{ih}^{(1)*}\right)$ denote the vector of pretest latent utilities of all items included in the optimal set of respondent $h$ at the pretest shopping trip $t$. Let $\ln U_{ih}^{(2)}$ denote the vector of posttest latent utilities of all items available to respondent $h$ at the posttest shopping trip $t$, and $\ln u\left(x_{ih}^{(2)*}\right)$ denote the vector of posttest latent utilities of all items included in the optimal set of respondent $h$ at the pretest shopping trip $t$.

(1) Generate $\ln u\left(x_{ih}^{(1)}\right)$, $\forall i = 1, 2, ..., K; b = 2, 3, ..., B; h = 1, 2, ..., H; t = 1, 2, ..., T_1$ and $\ln u\left(x_{ih}^{(2)}\right)$, $\forall i = 1, 2, ..., K; b = 1, 2, ..., B; h = 1, 2, ..., H; t = 1, 2, ..., T_2$. Then retain the error realizations.
Random-walk Metropolis-Hasting algorithm is used to generate the draw of $\beta_h^{(1)^*}$. Let $\beta_h^{(1)^*(o)}$ be the previous draw for $\beta_h^{(1)^*}$, and the next draw $\beta_h^{(1)^*(n)}$ is given by

$$\beta_h^{(1)^*(n)} = \beta_h^{(1)^*(o)} + \Delta \beta,$$

where $\Delta \beta$ is a draw from a candidate generating density Normal $(0, \sigma_{\beta_h^{(1)}}^2)$. Then, the candidate draw of $\beta_h^{(2)^*}$ is

$$\beta_h^{(2)^*(n)} = C_h \beta_h^{(1)^*(n)}.$$

Let $\ln u^{(o)}(x_{th}^{(1)^*})$ be the previous draw for log utility $\ln u(x_{th}^{(1)^*})$, and the next draw $\ln u^{(n)}(x_{th}^{(1)^*})$ is given by

$$\ln u^{(n)}(x_{th}^{(1)^*}) = z_{ht}^* \beta_h^{(1)^*(n)} + e_{ht}^*.$$
Let $\ln u^{(o)}(x_{th}^{(2)*})$ be the previous draw for log utility $\ln u(x_{th}^{(2)*})$, and the next draw $\ln u^{(n)}(x_{th}^{(2)*})$ is given by

$$
\ln u^{(n)}(x_{th}^{(2)*}) = z^*_n \beta_h^{(2)(n)} + e^*_n.
$$

The parameter $\sigma^2_{\beta_h^{(1)h}}$ of this density is adjusted to ensure 30% acceptance rate.

Thus, the probability of move $\beta^{(1)}_h$ is given by

$$
\kappa(\beta^{(1)(o)}_h, \beta^{(1)(n)}_h) = \min \left( \frac{\pi(\beta^{(1)(n)}_h)}{\pi(\beta^{(1)(o)}_h)}, 1 \right)
$$

and

$$
\frac{\pi(\beta^{(o)}_h)}{\pi(\beta^{(o)}_h)} = \frac{\exp \left\{ -\frac{1}{2} \sum_{t=1}^{T_1} \tilde{e}_{th}^{(1)*} \tilde{\Sigma}_{th}^{(1)*} \tilde{e}_{th}^{(1)*} \right\} \cdot \exp \left\{ -\frac{1}{2} \sum_{t=1}^{T_2} \tilde{e}_{th}^{(2)*} \tilde{\Sigma}_{th}^{(2)*} \tilde{e}_{th}^{(2)*} \right\}}{\exp \left\{ -\frac{1}{2} \sum_{t=1}^{T_1} \tilde{e}_{th}^{(1)*} \tilde{\Sigma}_{th}^{(1)*} \tilde{e}_{th}^{(1)*} \right\} \cdot \exp \left\{ -\frac{1}{2} \sum_{t=1}^{T_2} \tilde{e}_{th}^{(2)*} \tilde{\Sigma}_{th}^{(2)*} \tilde{e}_{th}^{(2)*} \right\}}.
$$

$$
= \frac{\exp \left\{ -\frac{1}{2} \left[ \tilde{\beta}^* - \tilde{\bar{\beta}} \right] \tilde{\nu}^{-1} \left[ \tilde{\beta}^* - \tilde{\bar{\beta}} \right] \right\} \cdot \prod_{t=1}^{T_1} I(\ln u^{(o)}(x_{th}^{(1)*})) \cdot \prod_{t=1}^{T_2} I(\ln u^{(n)}(x_{th}^{(2)*}))}{\exp \left\{ -\frac{1}{2} \left[ \tilde{\beta}^* - \tilde{\bar{\beta}} \right] \tilde{\nu}^{-1} \left[ \tilde{\beta}^* - \tilde{\bar{\beta}} \right] \right\} \cdot \prod_{t=1}^{T_1} I(\ln u^{(o)}(x_{th}^{(1)*})) \cdot \prod_{t=1}^{T_2} I(\ln u^{(n)}(x_{th}^{(2)*}))}
$$

$$
= \frac{\exp \left\{ -\frac{1}{2} \left[ \tilde{\beta}^* - \tilde{\bar{\beta}} \right] \tilde{\nu}^{-1} \left[ \tilde{\beta}^* - \tilde{\bar{\beta}} \right] \right\} \cdot \prod_{t=1}^{T_1} I(\ln u^{(o)}(x_{th}^{(1)*})) \cdot \prod_{t=1}^{T_2} I(\ln u^{(n)}(x_{th}^{(2)*}))}{\exp \left\{ -\frac{1}{2} \left[ \tilde{\beta}^* - \tilde{\bar{\beta}} \right] \tilde{\nu}^{-1} \left[ \tilde{\beta}^* - \tilde{\bar{\beta}} \right] \right\} \cdot \prod_{t=1}^{T_1} I(\ln u^{(o)}(x_{th}^{(1)*})) \cdot \prod_{t=1}^{T_2} I(\ln u^{(n)}(x_{th}^{(2)*}))}
$$

$$
= \prod_{t=1}^{T_1} I(\ln u^{(o)}(x_{th}^{(1)*})) \cdot \prod_{t=1}^{T_2} I(\ln u^{(n)}(x_{th}^{(2)*}))
$$
Random-walk Metropolis-Hasting algorithm is used to generate the draw of $T_h$. Let $T_h^{(o)}$ be the previous draw for $T_h$, and the next draw $T_h^{(n)}$ is given by

$$T_h^{(n)} = T_h^{(o)} + \Delta T,$$

where $\Delta T$ is a draw from a candidate generating density Normal $(0, 1)$. Let

$$\ln u^{(o)}(x_{th}^{(1)*})$$

be the previous draw for log utility $\ln u(x_{th}^{(1)*})$, and the next draw $\ln u^{(n)}(x_{th}^{(1)*})$ is given by

$$\ln u^{(n)}(x_{th}^{(1)*}) = z_{lh}^{(n)*} \beta_{h}^{(l)} + e_{lh}^{*}.$$
Let \( \ln u^{(o)}(x^{(2)*}_{th}) \) be the previous draw for log utility \( \ln u(x^{(2)*}_{th}) \), and the next draw \( \ln u^{(n)}(x^{(2)*}_{th}) \) is given by

\[
\ln u^{(n)}(x^{(2)*}_{ht}) = z^{(n)*}_{ht} \beta^{(2)}_{ht} + e^{*}_{ht}.
\]

The choice for parameters of this density ensures 30% acceptance rate. Thus, the probability of move \( T_h \) is given by

\[
\kappa(T^{(o)}_h, T^{(n)}_h) = \min \left( \frac{\pi(T^{(n)}_h)}{\pi(T^{(o)}_h)}, 1 \right)
\]

and

\[
\frac{\pi(T^{(n)}_h)}{\pi(T^{(o)}_h)} = \frac{\exp \left\{ -\frac{1}{2\sigma_T^2} \left[ T^{(n)}_h - \bar{T} \right]^2 \right\} \prod_{t=1}^{T_1} I(\ln u^{(o)}(x^{(1)*}_{th})) \cdot \prod_{t=1}^{T_2} I(\ln u^{(n)}(x^{(2)*}_{th}))}{\exp \left\{ -\frac{1}{2\sigma_T^2} \left[ T^{(o)}_h - \bar{T} \right]^2 \right\} \cdot \prod_{t=1}^{T_1} I(\ln u^{(o)}(x^{(1)*}_{th})) \cdot \prod_{t=1}^{T_2} I(\ln u^{(o)}(x^{(2)*}_{th}))}
\]
(4) Generate $\beta_{1,h}$ and $\psi_h$, $\forall h=1,2,\ldots,H$

Since $\beta_{01,h} = \beta_{1,h} + \gamma_h m_{1h} + \psi_h^\beta_{02,h}$, the full conditional of $\beta_{1,h}$ and $\psi_h$ is

$$
\begin{align*}
[\beta_{1,h}, \psi_h \mid \text{rest}] & \propto \prod_{t=1}^{T_h} I \left( \ln U_{ih}^{(2)} \right) \cdot \left[ \ln U_{ih}^{(2)} \mid \beta_{1,h}^{(2)} \right] \left[ \beta_{1,h} \mid \tilde{\beta}_{01}, \sigma_{01}^2 \right] \left[ \psi_h \mid \tilde{\psi}, \sigma_{\psi}^2 \right] \\
& \propto \prod_{t=1}^{T_h} I \left( \ln u \left( x_{ih}^{(2)} \right) \right) \cdot \ln u \left( x_{ih}^{(2)} - z_{ih}^{(2)} \beta_{1,h}^{(2)} \right) \left[ \ln u \left( x_{ih}^{(2)} - z_{ih}^{(2)} \beta_{1,h}^{(2)} \right) \right] \\
& \propto \exp \left\{ -\frac{1}{2} \sum_{t=1}^{T_h} \left[ \ln u \left( x_{ih}^{(2)} \right) \right] \right\} \cdot \exp \left\{ -\frac{1}{2\sigma_{01}^2} \left[ \beta_{1,h} - \tilde{\beta}_{01} \right]^2 \right\} \cdot \prod_{t=1}^{T_h} I \left( \ln u \left( x_{ih}^{(2)} \right) \right) \\
& \propto \exp \left\{ -\frac{1}{2} \sum_{t=1}^{T_h} \left[ \ln u \left( x_{ih}^{(2)} \right) \right] \right\} \cdot \exp \left\{ -\frac{1}{2\sigma_{01}^2} \left[ \beta_{1,h} - \tilde{\beta}_{01} \right]^2 \right\} \cdot \prod_{t=1}^{T_h} I \left( \ln u \left( x_{ih}^{(2)} \right) \right)
\end{align*}
$$

Random-walk Metropolis-Hasting algorithm is used to generate the draws of $\beta_{1,h}$ and $\psi_h$. Let $\beta_{1,h}^{(o)}$ be the previous draw for $\beta_{1,h}$, and the next draw $\beta_{1,h}^{(n)}$ is given by

$$
\beta_{1,h}^{(n)} = \beta_{1,h}^{(o)} + \Delta \beta,
$$

where $\Delta \beta$ is a draw from a candidate generating density Normal $(0, 0.3)$. Let $\psi_h^{(o)}$ be the previous draw for $\psi_h$, and the next draw $\psi_h^{(n)}$ is given by

$$
\psi_h^{(n)} = \psi_h^{(o)} + \Delta \psi,
$$

where $\Delta \psi$ is a draw from a candidate generating density Normal $(0, 0.3)$. Then, the candidate draw of $\beta_{01,h}^{(2)}$ is

$$
\beta_{01,h}^{(n)} = \beta_{1,h}^{(n)} + \gamma_h m_{1h} + \psi_h^{(n)} \beta_{02,h}^{(1)}
$$
Let $\ln u^{(o)}\left(x_{n}^{(2)*}\right)$ be the previous draw for log utility $\ln u\left(x_{n}^{(2)*}\right)$, and the next draw $\ln u^{(n)}\left(x_{n}^{(2)*}\right)$ is given by

$$\ln u^{(n)}\left(x_{n}^{(2)*}\right) = z_{n}^{*} \beta_{n}^{(2)(n)} + e_{n}^{*}.$$  

The choice for parameters of this density ensures 30% acceptance rate. Thus, the probability of move $\beta_{h}^{(1)}$ is given by

$$\kappa\left(\left\{ \beta_{h}^{(o)}, \psi_{h}^{(o)} \right\}, \left\{ \beta_{h}^{(n)}, \psi_{h}^{(n)} \right\}\right) = \min\left( \frac{\pi\left( \beta_{h}^{(n)}, \psi_{h}^{(n)} \right)}{\pi\left( \beta_{h}^{(o)}, \psi_{h}^{(o)} \right)}, 1 \right)$$

and

$$\frac{\pi\left( \beta_{h}^{(n)}, \psi_{h}^{(n)} \right)}{\pi\left( \beta_{h}^{(o)}, \psi_{h}^{(o)} \right)} = \left\{ \exp\left\{-\frac{1}{2\sigma_{o1}^{2}} \left[ \beta_{h}^{(n)} - \beta_{o1}^{(o)} \right]^{2}\right\}, \right. \left. \exp\left\{-\frac{1}{2\sigma_{o2}^{2}} \left[ \beta_{h}^{(n)} - \beta_{o2}^{(o)} \right]^{2}\right\}, \right. \left. \exp\left\{-\frac{1}{2\sigma_{\psi}^{2}} \left[ \psi_{h}^{(n)} - \psi^{(o)} \right]^{2}\right\}, \prod_{i=1}^{T} I\left( \ln u^{(n)} \left( x_{h}^{(2)*} \right) \right) \right\}.$$  

(5) Generate $\gamma_{h}, \forall h=1,2,..,H$ and $M_{h} \neq 0$

Since $\beta_{h}^{(1)} = \left( \beta_{2h}, \beta_{T_{h}} \right)^{(1)'}, \beta_{h}^{(2)} = \left( \beta_{01h}, \beta_{2h}, \beta_{T_{h}} \right)^{(2)'},$

$$\beta_{01h} = \beta_{1h} + \gamma_{h} m_{1h} + \psi_{h} \beta_{02h}^{(1)}, \beta_{2h} = C_{h} \beta_{2h}^{(1)},$$ and

$$C_{h} = diag\left(1 + \gamma_{h} m_{2h}, \cdots, 1 + \gamma_{h} m_{2h}, 1 + \theta_{1,h} M_{h}, \cdots, 1 + \theta_{nvar,h} M_{h}, 1 + \theta_{x} M_{h}\right),$$ the conditional distribution of $\gamma_{h}$ is

197
Random-walk Metroplis-Hasting algorithm is used to generate the draws of $\gamma_h$. Let $\gamma_h^{(o)}$ be the previous draw for $\gamma_h$, and the next draw $\gamma_h^{(n)}$ is given by

$$\gamma_h^{(n)} = \gamma_h^{(o)} + \Delta \gamma,$$

where $\Delta \gamma$ is a draw from a candidate generating density Normal $(0, \sigma^2_{\gamma_h})$. Then, the candidate draw of $\beta_{(2)}^{(2)}$ is

$$C_h^{(n)} = diag \left( 1 + \gamma_h^{(n)}m_{2h}, \ldots, 1 + \gamma_h^{(n)}m_{nh}, 1 + \theta_{1,h}M_h, \ldots, 1 + \theta_{nvar,h}M_h, 1 + \theta_{z,h}M_h \right)$$

$$\beta_{01,h}^{(n)} = \beta_{01,h} + \gamma_h^{(n)} + \psi_h \beta_{02,h}^{(1)}$$

$$\beta_{2h}^{(2)(n)} = C_h^{(n)} \beta_{2h}^{(1)}$$

Let $\ln u^{(n)}(x_{ih}^{(2)r})$ be the previous draw for log utility $\ln u(x_{ih}^{(2)r})$, and the next draw $\ln u^{(n)}(x_{ih}^{(2)r})$ is given by

$$\ln u^{(n)}(x_{ih}^{(2)r}) = z_{ih}^{*} \beta_{(2)(n)}^{2h} + e_{ih}^{*}.$$
\[ \kappa(\gamma^{(o)}_h, \gamma^{(n)}_h) = \min\left\{ \frac{\pi(\gamma^{(o)}_h)}{\pi(\gamma^{(n)}_h)}, 1 \right\} \]

and

\[ \frac{\pi(\gamma^{(n)}_h)}{\pi(\gamma^{(o)}_h)} = \exp\left\{ -\frac{1}{2} \left[ \gamma^{(n)}_h - \bar{\gamma} \right] V^{-1}_\gamma \left[ \gamma^{(n)}_h - \bar{\gamma} \right] \right\} \cdot \prod_{i=1}^{T_h} I\left( \ln u^{(n)}(x^{(2)_r}_{ih}) \right) \]

\[ \cdot \exp\left\{ -\frac{1}{2} \left[ \gamma^{(o)}_h - \bar{\gamma} \right] V^{-1}_\gamma \left[ \gamma^{(o)}_h - \bar{\gamma} \right] \right\} \cdot \prod_{i=1}^{T_h} I\left( \ln u^{(o)}(x^{(2)_r}_{ih}) \right) \]

(6) Generate \( \theta_{ih}, \forall i=1,2,\ldots,n_{var}; h=1,2,\ldots, H \) and \( M_h \neq 0 \)

Let \( \Theta_h = \left( \theta_{1h}, \ldots, \theta_{n_{var}h}, \theta_x \right) \).

Since \( \beta^{(1)}_h = \left( \beta_{2h}, \beta_{3h} \right)^{(1)_r}, \beta^{(2)}_h = \left( \beta_{01h}, \beta_{2h}, \beta_{2T_h} \right)^{(2)_r}, \beta^{(3)}_h = C_h \beta^{(1)}_h, \) and

\[ C_h = \text{diag} \left( 1 + \gamma_{h}m_{2h}, \ldots, 1 + \gamma_{h}m_{2h}, 1 + \theta_{1h}M_{h}, \ldots, 1 + \theta_{n_{var}h}M_{h}, 1 + \theta_xM_{h} \right), \]

the conditional distribution of \( \theta_{ih} \) is

\[
\left[ \Theta_h \mid \text{rest} \right] \propto \prod_{i=1}^{T_h} I\left( \ln U^{(2)}_{ih} \right) \left[ \ln U^{(2)}_{ih} \mid \beta^{(3)}_h \right] \prod_{i=1}^{n_{var}} \left[ \theta_{ih} \mid \bar{\theta}, V_{\theta} \right] \\
\propto \prod_{i=1}^{T_h} I\left( \ln u^{(2)_r}_{ih} \right) \left[ \ln u^{(2)_r}_{ih} \mid \beta^{(3)}_h \right] \prod_{i=1}^{n_{var}} \left[ \theta_{ih} \mid \bar{\theta}, V_{\theta} \right] \\
\propto \exp\left\{ -\frac{1}{2} \sum_{i=1}^{T_h} \left[ \ln u^{(2)_r}_{ih} \gamma^{(2)_r}_{ih} \right] \Sigma^{(2)_r}^{-1} \left[ \ln u^{(2)_r}_{ih} - z^{(2)_r}_{ih} \beta^{(3)}_h \right] \right\} \cdot \prod_{i=1}^{T_h} I\left( \ln u^{(2)_r}_{ih} \right) \\
\exp\left\{ -\frac{1}{2} \sum_{i=1}^{n_{var}} \left[ \theta_{ih} - \bar{\theta} \right] V^{-1}_\theta \left[ \theta_{ih} - \bar{\theta} \right] \right\} \cdot \prod_{i=1}^{T_h} I\left( \ln u^{(2)_r}_{ih} \right) \\
\propto \left[ \frac{1}{2} \sum_{i=1}^{T_h} \left[ \gamma^{(2)_r}_{ih} e^{(2)_r}_{ih} \right] e^{(2)_r}_{ih} \right] \cdot \prod_{i=1}^{T_h} I\left( \ln u^{(2)_r}_{ih} \right) \\
\exp\left\{ -\frac{1}{2} \sum_{i=1}^{n_{var}} \left[ \theta_{ih} - \bar{\theta} \right] V^{-1}_\theta \left[ \theta_{ih} - \bar{\theta} \right] \right\} \cdot \prod_{i=1}^{T_h} I\left( \ln u^{(2)_r}_{ih} \right) \\
\end{align*} \]
Random-walk Metropolis-Hastings algorithm is used to generate the draws of $\Theta_h$.

Let $\Theta_h^{(o)}$ be the previous draw for $\Theta_h$, and the next draw $\Theta_h^{(n)}$ is given by

$$\Theta_h^{(n)} = \Theta_h^{(o)} + \Delta \theta,$$

where $\Delta \theta$ is a draw from a candidate generating density Normal $(0, \sigma_{\theta h}^2)$. Then, the candidate draw of $\beta_h^{(2)}$ is

$$C_h^{(n)} = \text{diag} \left( 1 + \gamma_h m_{2h}, \ldots, 1 + \gamma_h m_{Bh}, 1 + \theta_{1h}^{(n)} M_{h}, \ldots, 1 + \theta_{x}^{(n)} M_{h} \right),$$

$$\beta_{T,h}^{(2)(n)} = \exp \left( \beta_{T,h}^{(2)(n)} \right) = \exp \left( \Theta_T^{(n)} M_{h} \rho_T^{(1)} \right),$$

$$\beta_{2h}^{(2)(n)} = C_h^{(n)} \beta_{2h}^{(1)}$$

$$\beta_h^{(2)} = (\beta_{01h}, \beta_{2h}^{(n)}, \beta_{T,h}^{(2)(n)})^{(2)(n)}.$$  

Let $\ln u^{(o)} (x_{ih}^{(2)(n)})$ be the previous draw for log utility $\ln u (x_{ih}^{(2)(n)})$, and the next draw $\ln u^{(n)} (x_{ih}^{(2)(n)})$ is given by

$$\ln u^{(n)} (x_{ih}^{(2)(n)}) = z_{ih}^{*} \beta_{h}^{(2)(n)} + e_{ih}^{*}.$$  

The parameter $\sigma_{\theta h}^2$ of this density is adjusted in each MCMC iteration to ensure 30\% acceptance rate. Thus, the probability of move $\beta_h^{(2)}$ is given by

$$\kappa \left( (\Theta_h^{(o)}, \Theta_h^{(n)}) \right) = \min \left( \frac{\pi \left( (\Theta_h^{(n)}) \right)}{\pi \left( (\Theta_h^{(o)}) \right)}, 1 \right)$$

and

$$\pi \left( (\Theta_h^{(o)}) \right) = \exp \left\{ -\frac{1}{2} \sum_{i=1}^{\text{num}} \left[ \theta_{i,h}^{(o)} - \overline{\theta} \right] V_0^{-1} \left[ \theta_{i,h}^{(o)} - \overline{\theta} \right] \right\} \cdot \prod_{t=1}^{T} I \left( \ln u^{(o)} (x_{ih}^{(2)(n)}) \right),$$

$$\pi \left( (\Theta_h^{(o)}) \right) = \exp \left\{ -\frac{1}{2} \sum_{i=1}^{\text{num}} \left[ \theta_{i,h}^{(o)} - \overline{\theta} \right] V_0^{-1} \left[ \theta_{i,h}^{(o)} - \overline{\theta} \right] \right\} \cdot \prod_{t=1}^{T} I \left( \ln u^{(o)} (x_{ih}^{(2)(n)}) \right).$$
(7) Generate $\bar{\beta}$

$$[\bar{\beta} \mid \text{rest}] \propto \prod_{h=1}^{H} [\beta_h \mid \bar{\beta}, V_\beta] \cdot [\bar{\beta} \mid \mu_0, V_0]$$

$$\sim N \left( H V_\beta^{-1} + V_0^{-1} \right)^{-1} \left[ V_\beta^{-1} \sum_{h=1}^{H} \beta_h + V_0^{-1} \mu_0 \right], \left[ H V_\beta^{-1} + V_0^{-1} \right]^{-1} \right)$$

(8) Generate $V_\beta$

$$[V_\beta \mid \text{rest}] \propto \prod_{h=1}^{H} [\beta_h \mid \bar{\beta}, V_\beta] \cdot [V_\beta \mid \nu, G]$$

$$\sim IW \left( G + \nu + n \text{var.} \sum_{h=1}^{H} (\beta_h - \bar{\beta})(\beta_h - \bar{\beta})' + G \right)$$

(9) Generate $T$

$$[T \mid \text{rest}] \propto \prod_{h=1}^{H} [T_h \mid \bar{T}, \sigma_T^2] \cdot [T \mid a, b^2]$$

$$\sim N \left( \frac{H}{\sigma_T^2} + \frac{1}{b^2} \right)^{-1} \left[ \frac{1}{b^2} + \frac{a}{\sigma_T^2} \right], \left( \frac{H}{\sigma_T^2} + \frac{1}{b^2} \right)^{-1} \right)$$

(10) Generate $\sigma_T^2$

$$[\sigma_T^2 \mid \text{rest}] \propto \prod_{h=1}^{H} [T_h \mid \bar{T}, \sigma_T^2] \cdot [\sigma_T^2 \mid \nu_T, s_T^2]$$

$$\sim \nu_T s_T^2 \chi^2_{\nu_T}, \text{where } \nu_T = H + \nu_T, \text{ and } \nu_T s_T^2 = \sum_{h=1}^{H} (T_h - \bar{T})^2 + \nu_T s_T^2$$

(11) Generate $\gamma$

$$[\gamma \mid \text{rest}] \propto \prod_{h=1}^{H} [\gamma_h \mid \bar{\gamma}, V_\gamma] \cdot [\bar{\gamma} \mid \mu_{\gamma_0}, V_{\gamma_0}]$$

$$\sim N \left( \left( H - \sum_{h=1}^{H} I(M_h = 1) \right) V_\gamma^{-1} + V_{\gamma_0}^{-1} \right)^{-1} \left[ V_\gamma^{-1} \sum_{h=1}^{H} \gamma_h + V_{\gamma_0}^{-1} \mu_\gamma \right], \left[ H - \sum_{h=1}^{H} I(M_h = 1) \right] V_\gamma^{-1} + V_{\gamma_0}^{-1} \right)^{-1}$$
(12) Generate $V_\gamma$

$$[V_\gamma | \text{rest}] \propto \prod_{h=1}^{H} \left[ \gamma_h | \bar{\gamma}, V_\gamma \right] \cdot [V_\gamma | \nu_\gamma, G_\gamma]$$

$$\sim IW \left( H - \sum_{h=1}^{H} I(M_h = 0) + \nu_\gamma, \sum_{h=1}^{H} (\gamma_h - \bar{\gamma})(\gamma_h - \bar{\gamma})' + G_\gamma \right)$$

(13) Generate $\bar{\theta}$

$$[\bar{\theta} | \text{rest}] \propto \prod_{h=1}^{H} \prod_{i=1}^{n_{\text{var}}} \left[ \bar{\theta}_{i,h} | \bar{\theta}, V_\theta \right] \cdot [\bar{\theta} | \mu_{\theta 0}, V_{\theta 0}]$$

$$\sim N \left( \left( H - \sum_{h=1}^{H} I(M_h = 1) \right) \cdot n_{\text{var}} \cdot V^{-1}_\theta + V^{-1}_{\theta 0}, V^{-1}_{\theta 0} \right)$$

(14) Generate $V_\theta$

$$[V_\theta | \text{rest}] \propto \prod_{h=1}^{H} \prod_{i=1}^{n_{\text{var}}} \left[ \bar{\theta}_{i,h} | \bar{\theta}, V_\theta \right] \cdot [V_\theta | \nu_\theta, G_\theta]$$

$$\sim IW \left( H - \sum_{h=1}^{H} I(M_h = 0) \right) \cdot n_{\text{var}} + \nu_\theta, \sum_{h=1}^{H} n_{\text{var}} \sum_{i=1}^{n_{\text{var}}} (\theta_{i,h} - \bar{\theta})(\theta_{i,h} - \bar{\theta})' + G_\theta$$
C.4. Additive Model

Pretest and posttest data are fit by the model specified in Equations (5.18), (5.20), (5.37)-(5.38). Let \( U_{th}^{(1)} \) denote the vector of pretest latent utilities of all items available to respondent \( h \) at the pretest shopping trip \( t \), and \( \ln u(x_{ih}^{(1)*}) \) denote the vector of pretest latent utilities of all items included in the optimal set of respondent \( h \) at the pretest shopping trip \( t \). Let \( U_{th}^{(2)} \) denote the vector of posttest latent utilities of all items available to respondent \( h \) at the posttest shopping trip \( t \), and \( \ln u(x_{ih}^{(2)*}) \) denote the vector of posttest latent utilities of all items included in the optimal set of respondent \( h \) at the pretest shopping trip \( t \).

1. Generate \( \ln u(x_{ih}^{(1)*}), \forall i = 1, 2, \ldots, K; b = 2, 3, \ldots, B; h = 1, 2, \ldots, H; t = 1, 2, \ldots, T_1 \) and \( \ln u(x_{ih}^{(2)*}), \forall i = 1, 2, \ldots, K; b = 1, 2, \ldots, B; h = 1, 2, \ldots, H; t = 1, 2, \ldots, T_2 \). Then retain the error realizations.
Random-walk Metropolis-Hasting algorithm is used to generate the draw of \( \beta_h^{(1)*} \).

Let \( \beta_h^{(1)* (o)} \) be the previous draw for \( \beta_h^{(1)*} \), and the next draw \( \beta_h^{(1)* (n)} \) is given by

\[
\beta_h^{(1)* (n)} = \beta_h^{(1)* (o)} + \Delta \beta ,
\]

where \( \Delta \beta \) is a draw from a candidate generating density Normal \((0, \sigma_{\beta_h}^2)\). Then, the candidate draw of \( \beta_h^{(2)*} \) is

\[
\beta_h^{(2)* (n)} = C_h \beta_h^{(1)* (n)}
\]

Let \( \ln u^{(o)}(x_h^{(1)*}) \) be the previous draw for log utility \( \ln u(x_h^{(1)*}) \), and the next draw \( \ln u^{(n)}(x_h^{(1)*}) \) is given by

\[
\ln u^{(n)}(x_h^{(1)*}) = z_h^* \beta_h^{(1)* (n)} + \epsilon_h^* .
\]
Let $\ln u^{(o)}(x_{th}^{(2)*})$ be the previous draw for log utility $\ln u(x_{th}^{(2)*})$, and the next draw $\ln u^{(n)}(x_{th}^{(2)*})$ is given by

$$\ln u^{(n)}(x_{th}^{(2)*}) = z_{th}^* \beta_{h}^{(2)(n)} + e_{th}^*. $$

The parameter $\sigma_{\beta h}^2$ of this density is adjusted in each MCMC iteration to ensure 30% acceptance rate. Thus, the probability of move $\beta_{h}^{(1)}$ is given by

$$\kappa(\beta_{h}^{(1)(o)}, \beta_{h}^{(1)(n)}) = \min \left\{ \frac{\pi(\beta_{h}^{(1)(n)})}{\pi(\beta_{h}^{(1)(o)})}, 1 \right\}$$

and

$$\frac{\pi(\beta_{h}^{(o)})}{\pi(\beta_{h}^{(o)})} = \frac{\exp \left\{ -\frac{1}{2} \sum_{t=1}^{T_1} \epsilon_{th}^{(1)*} \Sigma_{th}^{(1)} \epsilon_{th}^{(1)*} \right\} \cdot \exp \left\{ -\frac{1}{2} \sum_{t=1}^{T_2} \epsilon_{th}^{(2)*} \Sigma_{th}^{(2)} \epsilon_{th}^{(2)*} \right\} \cdot \exp \left\{ -\frac{1}{2} \left[ \beta_{h}^{(1)(o)} - \bar{\beta} \right] V_{\beta}^{-1} \left[ \beta_{h}^{(1)(o)} - \bar{\beta} \right]^\top \right\} \cdot \prod_{t=1}^{T_1} I \left( \ln u^{(o)}(x_{th}^{(1)*}) \right) \cdot \prod_{t=1}^{T_2} I \left( \ln u^{(o)}(x_{th}^{(2)*}) \right)}{\exp \left\{ -\frac{1}{2} \sum_{t=1}^{T_1} \epsilon_{th}^{(1)*} \Sigma_{th}^{(1)} \epsilon_{th}^{(1)*} \right\} \cdot \exp \left\{ -\frac{1}{2} \sum_{t=1}^{T_2} \epsilon_{th}^{(2)*} \Sigma_{th}^{(2)} \epsilon_{th}^{(2)*} \right\} \cdot \exp \left\{ -\frac{1}{2} \left[ \beta_{h}^{(1)(o)} - \bar{\beta} \right] V_{\beta}^{-1} \left[ \beta_{h}^{(1)(o)} - \bar{\beta} \right]^\top \right\} \cdot \prod_{t=1}^{T_1} I \left( \ln u^{(o)}(x_{th}^{(1)*}) \right) \cdot \prod_{t=1}^{T_2} I \left( \ln u^{(o)}(x_{th}^{(2)*}) \right)}.$$
(3) Generate $T_h, \forall h = 1, 2, \ldots, H$

$$[T_h \mid \text{rest}] \propto \prod_{t=1}^{T_h} \int \left[ \ln U_{th}^{(1)} \right] \cdot \left[ \ln U_{th}^{(1)} \mid \beta_h^{(1)} \right].$$

$$\prod_{t=1}^{T_h} \left[ \ln U_{th}^{(2)} \mid \ln U_{th}^{(2)} \mid \beta_h^{(2)} \right] \cdot \left[ T_h \mid \bar{T}, \sigma_T^2 \right]$$

$$\propto \prod_{t=1}^{T_h} \int \left[ u \left( x_{th}^{(1)*} \right) \right] \cdot \left[ u \left( x_{th}^{(1)*} \right) \mid \beta_h^{(1)} \right].$$

$$\prod_{t=1}^{T_h} \int \left[ u \left( x_{th}^{(2)*} \right) \right] \cdot \left[ y_{th}^{(2)} \mid u \left( x_{th}^{(2)*} \right) \right] \cdot \left[ T_h \mid \bar{T}, \sigma_T^2 \right]$$

$$\propto \exp \left\{ -\frac{1}{2} \sum_{t=1}^{T_h} \left[ \ln u \left( x_{th}^{(1)*} \right) - z_{th}^{(1)*} \beta_h^{(1)} \right]^T \Sigma^{(1)^{-1}} \left[ \ln u \left( x_{th}^{(1)*} \right) - z_{th}^{(1)*} \beta_h^{(1)} \right] \right\}.$$  

$$\exp \left\{ -\frac{1}{2} \sum_{t=1}^{T_h} \left[ \ln u \left( x_{th}^{(2)*} \right) - z_{th}^{(2)*} \beta_h^{(2)} \right]^T \Sigma^{(2)^{-1}} \left[ \ln u \left( x_{th}^{(2)*} \right) - z_{th}^{(2)*} \beta_h^{(2)} \right] \right\}.$$  

$$\exp \left\{ -\frac{1}{2} \sum_{t=1}^{T_h} \left[ \ln u \left( x_{th}^{(1)*} \right) - z_{th}^{(1)*} \beta_h^{(1)} \right]^T \Sigma^{(1)^{-1}} \left[ \ln u \left( x_{th}^{(1)*} \right) - z_{th}^{(1)*} \beta_h^{(1)} \right] \right\}.$$  

$$\exp \left\{ -\frac{1}{2} \sum_{t=1}^{T_h} \left[ \ln u \left( x_{th}^{(2)*} \right) - z_{th}^{(2)*} \beta_h^{(2)} \right]^T \Sigma^{(2)^{-1}} \left[ \ln u \left( x_{th}^{(2)*} \right) - z_{th}^{(2)*} \beta_h^{(2)} \right] \right\}.$$  

$$\exp \left\{ -\frac{1}{2} \sum_{t=1}^{T_h} \left[ \ln u \left( x_{th}^{(1)*} \right) - z_{th}^{(1)*} \beta_h^{(1)} \right]^T \Sigma^{(1)^{-1}} \left[ \ln u \left( x_{th}^{(1)*} \right) - z_{th}^{(1)*} \beta_h^{(1)} \right] \right\}.$$  

Random-walk Metroplis-Hasting algorithm is used to generate the draw of $T_h$. Let $T_h^{(o)}$ be the previous draw for $T_h$, and the next draw $T_h^{(n)}$ is given by

$$T_h^{(n)} = T_h^{(o)} + \Delta T,$$

where $\Delta T$ is a draw from a candidate generating density Normal $(0, 1)$. Let

$\ln u^{(o)} \left( x_{th}^{(1)*} \right)$ be the previous draw for log utility $\ln u \left( x_{th}^{(1)*} \right)$, and the next draw

$\ln u^{(n)} \left( x_{th}^{(1)*} \right)$ is given by

$$\ln u^{(n)} \left( x_{th}^{(1)*} \right) = z_{th}^{(n)*} \beta_h^{(1)} + e_{th}^*.$$
Let \( \ln u^{(o)}(x_{th}^{(2)*}) \) be the previous draw for log utility \( \ln u(x_{th}^{(2)*}) \), and the next draw \( \ln u^{(n)}(x_{th}^{(2)*}) \) is given by

\[
\ln u^{(n)}(x_{th}^{(2)*}) = z_{th}^{(n)*} \beta_{th}^{(2)} + e_{th}^{*}.
\]

The choice for parameters of this density ensures 30% acceptance rate. Thus, the probability of move \( T_h \) is given by

\[
\kappa(T_h^{(o)}, T_h^{(n)}) = \min \left( \frac{\pi(T_h^{(n)})}{\pi(T_h^{(o)})}, 1 \right)
\]

and

\[
\frac{\pi(T_h^{(n)})}{\pi(T_h^{(o)})} = \exp \left\{ -\frac{1}{2\sigma^2} \left[ T_h^{(n)} - T_h^{(o)} \right]^2 \right\} \prod_{t=1}^{T} \left( \ln u^{(o)}(x_{th}^{(1)*}) \right) \cdot \prod_{t=1}^{T} \left( \ln u^{(n)}(x_{th}^{(2)*}) \right)
\]

\[
\frac{\pi(T_h^{(o)})}{\pi(T_h^{(n)})} = \exp \left\{ -\frac{1}{2\sigma^2} \left[ T_h^{(o)} - T_h^{(n)} \right]^2 \right\} \cdot \prod_{t=1}^{T} \left( \ln u^{(o)}(x_{th}^{(1)*}) \right) \cdot \prod_{t=1}^{T} \left( \ln u^{(n)}(x_{th}^{(2)*}) \right)
\]
(4) Generate $\beta_{1,h}$ and $\psi_h$, $\forall h=1,2,\ldots,H$

Since $\beta_{01,h} = \beta_{1,h} + \gamma_h m_{1h} + \psi_h \beta_{02,h}^{(1)}$, the full conditional of $\beta_{1,h}$ and $\psi_h$ is

$$
[\beta_{1h}, \psi_h | \text{rest}] \propto \prod_{t=1}^{T} I\left(\ln U_{th}^{(2)}\right) \cdot [\ln U_{th}^{(2)} | \beta_{2h}^{(2)}] \left[ \beta_{1h} | \beta_{01h}, \sigma_{01}^2 \right] \left[ \psi_h | \psi, \sigma_{\psi}^2 \right] \\
= \prod_{t=1}^{T} \left( \ln u\left(x_{th}^{(2)}\right) \right) \cdot [\ln u\left(x_{th}^{(2)}\right) | \beta_{2h}^{(2)}] \left[ \beta_{1h} | \beta_{01h}, \sigma_{01}^2 \right] \left[ \psi_h | \psi, \sigma_{\psi}^2 \right] \\
= \exp \left\{ -\frac{1}{2} \sum_{t=1}^{T} \left[ \ln u\left(x_{th}^{(2)}\right) - \beta_{2h}^{(2)} \left[ \Sigma^{(2)} \right]^{-1} \left[ \ln u\left(x_{th}^{(2)}\right) - \beta_{2h}^{(2)} \right] \right] \right\} \cdot \exp \left\{ -\frac{1}{2 \sigma_{01}^2} \left[ \beta_{1h} - \bar{\beta}_{01h} \right]^2 \right\} \cdot \frac{1}{T} \prod_{t=1}^{T} I\left(\ln u\left(x_{th}^{(2)}\right)\right) \\
= \exp \left\{ -\frac{1}{2} \sum_{t=1}^{T} e_{th}^{(2)} \Sigma^{(2)} e_{th}^{(2)} \right\} \cdot \exp \left\{ -\frac{1}{2 \sigma_{01}^2} \left[ \beta_{1h} - \bar{\beta}_{01h} \right]^2 \right\} \cdot \frac{1}{T} \prod_{t=1}^{T} I\left(\ln u\left(x_{th}^{(2)}\right)\right) \\
= \exp \left\{ -\frac{1}{2 \sigma_{\psi}^2} \left[ \psi_h - \bar{\psi} \right]^2 \right\} \cdot \prod_{t=1}^{T} I\left(\ln u\left(x_{th}^{(2)}\right)\right)
$$

Random-walk Metroplis-Hasting algorithm is used to generate the draws of $\beta_{1,h}$ and $\psi_h$. Let $\beta_{1,h}^{(o)}$ be the previous draw for $\beta_{1,h}$, and the next draw $\beta_{1,h}^{(n)}$ is given by

$$
\beta_{1,h}^{(n)} = \beta_{1,h}^{(o)} + \Delta \beta,
$$

where $\Delta \beta$ is a draw from a candidate generating density Normal $(0, 0.3)$. Let $\psi_h^{(o)}$ be the previous draw for $\psi_h$, and the next draw $\psi_h^{(n)}$ is given by

$$
\psi_h^{(n)} = \psi_h^{(o)} + \Delta \psi,
$$

where $\Delta \psi$ is a draw from a candidate generating density Normal $(0, 0.3)$. Then, the candidate draw of $\beta_{01,h}^{(2)}$ is

$$
\beta_{01,h}^{(n)} = \beta_{1,h}^{(n)} + \gamma_h m_{1h} + \psi_h^{(n)} \beta_{02,h}^{(1)}
$$
Let \( \ln u^{(o)}(x_{th}^{(2)*}) \) be the previous draw for log utility \( \ln u(x_{th}^{(2)*}) \), and the next draw \( \ln u^{(n)}(x_{th}^{(2)*}) \) is given by

\[
\ln u^{(n)}(x_{th}^{(2)*}) = z_{th}^* \beta_h^{(2)(n)} + e_{th}^*.
\]

The choice for parameters of this density ensures 30% acceptance rate. Thus, the probability of move \( \beta^{(1)}_h \) is given by

\[
\kappa \left( \{ \beta^{(o)}_h, \psi_h^{(o)} \}, \{ \beta^{(n)}_h, \psi_h^{(n)} \} \right) = \min \left( \frac{\pi \left( \beta^{(n)}_h, \psi_h^{(n)} \right)}{\pi \left( \beta^{(o)}_h, \psi_h^{(o)} \right)}, 1 \right)
\]

and

\[
\pi \left( \beta^{(o)}_h, \psi_h^{(o)} \right) \frac{\exp \left\{ \frac{-1}{2\sigma_{\psi}^2} \left[ \psi_h^{(o)} - \psi \right]^2 \right\}}{\exp \left\{ \frac{-1}{2\sigma_{\psi}^2} \left[ \psi_h^{(n)} - \psi \right]^2 \right\}} \cdot \prod_{r=1}^{T} I \left( \ln u^{(n)}(x_{th}^{(2)*}) \right)
\]

\[
\pi \left( \beta^{(n)}_h, \psi_h^{(n)} \right) \frac{\exp \left\{ \frac{-1}{2\sigma_{\psi}^2} \left[ \psi_h^{(n)} - \psi \right]^2 \right\}}{\exp \left\{ \frac{-1}{2\sigma_{\psi}^2} \left[ \psi_h^{(o)} - \psi \right]^2 \right\}} \cdot \prod_{r=1}^{T} I \left( \ln u^{(o)}(x_{th}^{(2)*}) \right)
\]

(5) Generate \( \gamma_h, \forall h=1,2,\ldots,H \) and \( M_h \neq 0 \)

Since \( \beta^{(1)}_h = (\beta_{2h}, \beta_{T,h})^{(1)'} \), \( \beta^{(2)}_h = (\beta_{01h}, \beta_{2h}, \beta_{T,h})^{(2)'} \),

\[
\beta_{01h} = \beta_{1h} + \gamma_h m_{1h} + \psi_h \beta_{02h}^{(1)}, \beta_{2h} = \beta_{2h} + \Delta_h, \text{ and}
\]

\[
\Delta_h = \left( \gamma_h m_{2h}, \ldots, \gamma_h m_{B_h}, \theta_{in} M_h, \ldots, \theta_{nivar_h} M_h, \theta_x M_h \right), \text{ the conditional distribution of } \gamma_h \text{ is}
\]

209
Random-walk Metroplis-Hasting algorithm is used to generate the draws of $\gamma_h$. Let $\gamma^{(o)}_h$ be the previous draw for $\gamma_h$, and the next draw $\gamma^{(n)}_h$ is given by

$$\gamma^{(n)}_h = \gamma^{(o)}_h + \Delta \gamma,$$

where $\Delta \gamma$ is a draw from a candidate generating density Normal (0, $\sigma^2_{\gamma_h}$). Then, the candidate draw of $\beta^{(2)}_h$ is

$$\Delta^{(n)}_h = \left( \gamma^{(n)}_{2h}; \ldots, \gamma^{(n)}_{mth}, \beta^{(1)}_{1,1h}, \ldots, \beta_{\text{var},, M, \theta, M} \right)$$

$$\beta^{(n)}_{01,1h} = \beta^{(1)}_{1,1h} + \gamma^{(n)}_{1,1h} + \psi_{h} \beta^{(3)}_{02,1h}$$

$$\beta^{(2,n)}_{2h} = \beta^{(1)}_{2h} + \Delta^{(n)}_h$$

Let $\ln u^{(o)}(x^{(2r)}_{2h})$ be the previous draw for log utility $\ln u(x^{(2r)}_{2h})$, and the next draw $\ln u^{(n)}(x^{(2r)}_{2h})$ is given by

$$\ln u^{(n)}(x^{(2r)}_{2h}) = z^{*}_{2h} \beta^{(2,n)}_{2h} + e^{*}_{2h}.$$ 

The parameter $\sigma^2_{\gamma_h}$ of this density is adjusted to ensure 30% acceptance rate. Thus, the probability of move $\beta^{(1)}_{1h}$ is given by
\[ \kappa(\gamma_h^{(o)}, \gamma_h^{(n)}) = \min \left( \frac{\pi(\gamma_h^{(o)})}{\pi(\gamma_h^{(n)})}, 1 \right) \]

and

\[
\begin{align*}
\pi(\gamma_h^{(n)}) &= \exp \left\{ -\frac{1}{2} \left[ \gamma_h^{(n)} - \overline{\gamma} \right] \Sigma_h^{-1} \left[ \gamma_h^{(n)} - \overline{\gamma} \right] \right\} \cdot \prod_{i=1}^{T_h} I\left( \ln u^{(o)}(x_{ih}^{(2)^r}) \right) \\
\pi(\gamma_h^{(o)}) &= \exp \left\{ -\frac{1}{2} \left[ \gamma_h^{(o)} - \overline{\gamma} \right] \Sigma_h^{-1} \left[ \gamma_h^{(o)} - \overline{\gamma} \right] \right\} \cdot \prod_{i=1}^{T_h} I\left( \ln u^{(o)}(x_{ih}^{(2)^r}) \right)
\end{align*}
\]

(6) Generate \( \theta_{ih}, \forall i=1,2,\ldots,nvar; h=1,2,\ldots, H \) and \( M_h \neq 0 \)

Let \( \Theta_h = (\theta_{1h}, \ldots, \theta_{nvarh}, \theta_x) \).

Since \( \beta_h^{(1)} = (\beta_{2h}, \beta_{3h})^{(1)^r}, \beta_h^{(2)} = (\beta_{01h}, \beta_{2h}, \beta_{T,h})^{(2)^r}, \beta_{2h}^{(2)} = \beta_{2h}^{(1)} + \Delta_h \), and

\[ \Delta_h = (\gamma_h m_{2h}, \ldots, \gamma_h m_{Nh}, \theta_{ih} M_h, \ldots, \theta_{nvarh} M_h, \theta_x M_h) \]

the conditional distribution of \( \theta_{ih} \) is

\[
\begin{align*}
[\Theta_h | \text{rest}] \propto & \prod_{i=1}^{T_h} I\left( \ln U_{ih}^{(2)} \right) \left[ \ln U_{ih}^{(2)} \mid \beta_{h}^{(2)} \right] \prod_{i=1}^{nvar} \left[ \theta_{ih} \mid \overline{\theta}, V_{\theta} \right] \\
& \propto \prod_{i=1}^{T_h} I\left( \ln u(x_{ih}^{(2)^r}) \right) \left[ \ln u(x_{ih}^{(2)^r}) \mid \beta_{h}^{(2)} \right] \prod_{i=1}^{nvar} \left[ \theta_{ih} \mid \overline{\theta}, V_{\theta} \right] \\
& \propto \exp \left\{ -\frac{1}{2} \sum_{i=1}^{nvar} \left[ \ln u(x_{ih}^{(2)^r}) - z_{ih}^{(2)^r} \beta_{h}^{(2)} \right] \Sigma_h^{-1} \left[ \ln u(x_{ih}^{(2)^r}) - z_{ih}^{(2)^r} \beta_{h}^{(2)} \right] \right\}. \\
& \exp \left\{ -\frac{1}{2} \sum_{i=1}^{nvar} \left[ \theta_{ih} - \overline{\theta} \right] V_{\theta}^{-1} \left[ \theta_{ih} - \overline{\theta} \right] \right\} \cdot \prod_{i=1}^{T_h} I\left( \ln u(x_{ih}^{(2)^r}) \right) \\
& \propto \exp \left\{ -\frac{1}{2} \sum_{i=1}^{nvar} \left[ \theta_{ih} - \overline{\theta} \right] V_{\theta}^{-1} \left[ \theta_{ih} - \overline{\theta} \right] \right\} \cdot \prod_{i=1}^{T_h} I\left( \ln u(x_{ih}^{(2)^r}) \right)
\end{align*}
\]
Random-walk Metropolis-Hasting algorithm is used to generate the draws of $\Theta_h$.

Let $\Theta_h^{(o)}$ be the previous draw for $\Theta_h$, and the next draw $\Theta_h^{(n)}$ is given by

$$\Theta_h^{(n)} = \Theta_h^{(o)} + \Delta \theta,$$

where $\Delta \theta$ is a draw from a candidate generating density $\text{Normal}(0, \sigma_{\theta h}^2)$. Then, the candidate draw of $\beta^{(2)}_h$ is

$$\Delta_h^{(o)} = \left( \gamma_h m_{2h}, \cdots, \gamma_h m_{wh}, \theta^{(n)}_{1h}, \cdots, \theta^{(n)}_{x}, \theta^{(n)}_{\text{inv},h}, \theta^{(n)}_x, \theta^{(n)}_x \right)$$

$$\beta^{(2)}_{1h} = \exp\left( \beta^{(2)}_{1h}^{(o)} \right) = \exp\left( \beta^{(1)}_{1h} + \theta^{(n)}_{1h} M_h \right)$$

$$\beta^{(2)}_{2h} = \beta^{(1)}_{2h} + \Delta_h^{(o)}$$

$$\beta^{(2)}_h = \left( \beta^{(n)}_{0h}, \beta^{(n)}_{2h}, \beta^{(n)}_{T,h} \right)^{(2)}$$

Let $\ln u^{(o)}(x^{(2)}_{ih})$ be the previous draw for log utility $\ln u(x^{(2)}_{ih})$, and the next draw $\ln u^{(n)}(x^{(2)}_{ih})$ is given by

$$\ln u^{(n)}(x^{(2)}_{ih}) = z^{*}_{ih} \beta^{(2)(n)}_h + e^{*}_{ih}.$$ 

The parameter $\sigma^{2}_{\theta h}$ of this density is adjusted in each MCMC iteration to ensure 30% acceptance rate. Thus, the probability of move $\beta^{(2)}_h$ is given by

$$\kappa(\Theta_h^{(o)}, \Theta_h^{(n)}) = \min\left( \frac{\pi(\Theta_h^{(n)})}{\pi(\Theta_h^{(o)})}, 1 \right)$$

and

$$\pi(\Theta_h^{(o)}) = \exp\left\{ -\frac{1}{2} \sum_{i=1}^{n_{\text{vars}}} \left[ \theta^{(o)}_{ih} - \bar{\theta} \right]^T V_{\theta}^{-1} \left[ \theta^{(o)}_{ih} - \bar{\theta} \right] \right\} \prod_{t=1}^{T} I\left( \ln u^{(o)}(x^{(2)}_{ih}) \right)$$

$$\pi(\Theta_h^{(o)}) = \exp\left\{ -\frac{1}{2} \sum_{i=1}^{n_{\text{vars}}} \left[ \theta^{(o)}_{ih} - \bar{\theta} \right]^T V_{\theta}^{-1} \left[ \theta^{(o)}_{ih} - \bar{\theta} \right] \right\} \prod_{t=1}^{T} I\left( \ln u^{(o)}(x^{(2)}_{ih}) \right)$$
(7) Generate $\beta$

$$\begin{align*}
[\beta | \text{rest}] &\propto \prod_{h=1}^{H} [\beta_h | \beta, V_\beta] \cdot [\beta | \mu_0, V_0] \\
&\sim \mathcal{N} \left( HV_\beta^{-1} + V_0^{-1} \right) \left[ V_\beta^{-1} \sum_{h=1}^{H} \beta_h + V_0^{-1} \mu_0 \right] \left( HV_\beta^{-1} + V_0^{-1} \right)^{-1}
\end{align*}$$

(8) Generate $V_\beta$

$$\begin{align*}
[V_\beta | \text{rest}] &\propto \prod_{h=1}^{H} [\beta_h | \beta, V_\beta] \cdot [V_\beta | \nu, G] \\
&\sim \mathcal{I}W \left( G + \nu + \text{var} \sum_{h=1}^{H} (\beta_h - \bar{\beta}) (\beta_h - \bar{\beta})' + G \right)
\end{align*}$$

(9) Generate $\tilde{T}$

$$\begin{align*}
[\tilde{T} | \text{rest}] &\propto \prod_{h=1}^{H} [T_h | \tilde{T}, \sigma_T^2] \cdot [\tilde{T} | a, b^2] \\
&\sim \mathcal{N} \left( \frac{H}{\sigma_T^2 + \frac{1}{b^2}} \right) \left( \sum_{h=1}^{H} \frac{T_h}{\sigma_T^2 + \frac{1}{b^2}} \right) \left( \frac{H}{\sigma_T^2 + \frac{1}{b^2}} \right)^{-1}
\end{align*}$$

(10) Generate $\sigma_T^2$

$$\begin{align*}
[\sigma_T^2 | \text{rest}] &\propto \prod_{h=1}^{H} [T_h | \tilde{T}, \sigma_T^2] \cdot [\sigma_T^2 | \nu_{T_0}, s_{T_0}^2] \\
&\sim \frac{\nu_{T_0} s_{T_0}^2}{\chi_{v_t}^2}, \text{where } v_t = H + \nu_{T_0} \text{ and } \nu_{T_0} s_{T_0}^2 = \sum_{h=1}^{H} (T_h - \bar{T})^2 + \nu_{T_0} s_{T_0}^2
\end{align*}$$

(11) Generate $\bar{\gamma}$

$$\begin{align*}
[\bar{\gamma} | \text{rest}] &\propto \prod_{h=1}^{H} \prod_{M_h \neq 0} [\gamma_h | \bar{\gamma}, V_\gamma] \cdot [\bar{\gamma} | \mu_{\gamma_0}, V_{\gamma_0}] \\
&\sim \mathcal{N} \left( \left( H - \sum_{h=1}^{H} I(M_h = 1) \right) \cdot V_\gamma^{-1} + V_{\gamma_0}^{-1} \right)^{-1} \left[ V_{\gamma_0}^{-1} \sum_{h=1}^{H} \gamma_h + V_{\gamma_0}^{-1} \mu_{\gamma_0} \right], \\
&\left( H - \sum_{h=1}^{H} I(M_h = 1) \right) \cdot V_\gamma^{-1} + V_{\gamma_0}^{-1}
\end{align*}$$
(12) Generate $V_\gamma$

$$\left[ V_\gamma \mid \text{rest} \right] \propto \prod_{h=1}^{H} \sum_{M_h \neq 0} \left[ \gamma_h \mid \overline{\gamma}, V_\gamma \right] \cdot \left[ V_\gamma \mid \nu_\gamma, G_\gamma \right]$$

$$\sim IW \left( H - \sum_{h=1}^{H} I(M_h = 0) + \nu_\gamma \sum_{h=1}^{H} \sum_{M_h \neq 0} (\gamma_h - \overline{\gamma}) (\gamma_h - \overline{\gamma})' + G_\gamma \right)$$

(13) Generate $\bar{\theta}$

$$\left[ \bar{\theta} \mid \text{rest} \right] \propto \prod_{h=1}^{H} \prod_{M_h \neq 0} \sum_{i=1}^{n_{\text{var}}} \left[ \theta_{i,h} \mid \bar{\theta}, V_\theta \right] \cdot \left[ \bar{\theta} \mid \mu_{\theta 0}, V_{\theta 0} \right]$$

$$\sim N \left( \left[ H - \sum_{h=1}^{H} I(M_h = 1) \cdot n_{\text{var}} \cdot V_\theta^{-1} + V_{\theta 0}^{-1} \right]^{-1} V_\theta^{-1} \sum_{h=1}^{H} \sum_{M_h \neq 0} \theta_{i,h} + V_{\theta 0}^{-1} \mu_{\theta 0} \right),$$

$$\left( \left[ H - \sum_{h=1}^{H} I(M_h = 1) \cdot n_{\text{var}} \cdot V_\theta^{-1} + V_{\theta 0}^{-1} \right]^{-1} \right)$$

(14) Generate $V_\theta$

$$\left[ V_\theta \mid \text{rest} \right] \propto \prod_{h=1}^{H} \prod_{M_h \neq 0} \sum_{i=1}^{n_{\text{var}}} \left[ \theta_{i,h} \mid \bar{\theta}, V_\theta \right] \cdot \left[ V_\theta \mid \nu_\theta, G_\theta \right]$$

$$\sim IW \left( H - \sum_{h=1}^{H} I(M_h = 0) \cdot n_{\text{var}} + \nu_\theta, \sum_{h=1}^{H} \sum_{M_h \neq 0} (\theta_{i,h} - \bar{\theta}) (\theta_{i,h} - \bar{\theta})' + G_\theta \right)$$
LIST OF REFERENCES


219