ESSAYS ON HETEROGENEITY IN CHOICE MODELING

By

Kwangpil Chang

B. A. (Spanish) Hankuk University of Foreign Studies
M. Sc. (Advertising) University of Illinois at Urbana-Champaign

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
Doctor of Philosophy

in
THE FACULTY OF GRADUATE STUDIES
FACULTY OF COMMERCE AND BUSINESS ADMINISTRATION

We accept this thesis as conforming
to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA
August 1998
© Kwangpil Chang, 1998
In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Faculty of Commerce and Business Administration

The University of British Columbia

2075 Wesbrook Place

Vancouver, Canada

V6T 1Z1

Date: Sept 14, 1998
Abstract

This thesis includes three essays which examine the implications of incorporating parameter heterogeneity, consideration set heterogeneity, and decision rule heterogeneity, respectively, in brand choice models.

In the first essay, we identify the conditions under which unaccounted for price response heterogeneity results in a spurious sticker shock effect. We show, using an analytical derivation, a simulation study and an empirical application to scanner panel data, that estimates of the sticker shock effect may be biased if households that are price sensitive in their brand choice decision are also more likely to respond to category marketing activity in their purchase timing decision.

The empirical results, from two product categories, show that the sticker shock coefficient from a Hierarchical Bayes model (which continuously accounts for price response heterogeneity) is statistically insignificant, providing no evidence of the existence of a sticker shock effect. In contrast, the corresponding coefficient from the standard model, which ignores this heterogeneity, is highly significant and supports the existence of a sticker shock effect. A posterior analysis of household parameters confirms the hypothesized relationship between price sensitivity in brand choice and responsiveness to promotional activity in purchase incidence, and is consistent with our explanation of the underlying cause of the bias in the standard model.

The second essay develops a new consideration set model that can be estimated with scanner panel data. In contrast to many previous approaches, which require enumeration of all possible consideration sets, we directly model uncertainty about including a brand in the consideration set. The resulting inclusion probabilities for brands reflect a “fuzzy”
consideration set in the sense that a brand belongs to the consideration set only probabilistically. The proposed fuzzy set model outperforms several previous consideration set models in two product categories (yogurt and ketchup).

We then apply the fuzzy set approach to examine the role of the consideration set in moderating the impact of advertising on price sensitivity. In contrast to the experimental findings of Mitra and Lynch (1995), we find no positive relationship between consideration set size and price sensitivity. Further empirical test may be necessary to confirm the hypothesized relationship.

In the third essay, we investigate the role of decision rule heterogeneity in brand choice behavior. We develop a flexible model, which allows for the uncertainty in decision rules used by the consumer. Specifically, we develop a Hierarchical Bayes model of reference price effects that accommodates both the sticker shock and reference-dependent formulations. In addition, we also incorporate the possibility that consumers may mix the two decision rules probabilistically. Therefore, the proposed model allows for three different decision hierarchies which incorporate sticker shock, reference-dependent and mixed rules respectively.

The empirical results show that consumers differ not only in their preference and response but also in their decision rules. On average, half the sample households appear to show loss aversion, i.e., follow a reference-dependent decision rule, while the remaining households do not seem to respond to reference prices. The proposed model provides a richer description of consumer choice processes than the comparison models that allow for only one model structure and ignore model uncertainty.
# Table of Contents

Abstract  

List of Tables  

List of Figures  

Acknowledgement  

1 Introduction  

1.1 Overview  

2 Response Heterogeneity and Spurious Sticker Shock Effects  

2.1 Overview  

2.2 Theory  

2.2.1 Description of the True Model  

2.2.2 Description of the Misspecified Model  

2.2.3 Analysis of the Misspecified Model  

2.3 Simulation Study  

2.3.1 Data Generation  

2.3.2 Predictions and Estimated Models  

2.3.3 Simulation Results  

2.4 Empirical Application  

2.4.1 Data  

2.4.2 Models Estimated
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>B.2.1</td>
<td>Generating $\rho^{h(m+1)}$</td>
<td>126</td>
</tr>
<tr>
<td>B.2.2</td>
<td>Generating $\mu^{(m+1)}$</td>
<td>127</td>
</tr>
<tr>
<td>B.2.3</td>
<td>Generating $\Sigma^{-1(m+1)}$</td>
<td>128</td>
</tr>
<tr>
<td>C.1</td>
<td>Priors and Pseudopriors</td>
<td>129</td>
</tr>
<tr>
<td>C.1.1</td>
<td>Priors on Model Probabilities</td>
<td>129</td>
</tr>
<tr>
<td>C.1.2</td>
<td>Priors on $\beta^h_s \mid M = s$</td>
<td>129</td>
</tr>
<tr>
<td>C.1.3</td>
<td>Pseudopriors on $\beta^h_s \mid M \neq s$</td>
<td>130</td>
</tr>
<tr>
<td>C.2</td>
<td>Full Conditionals and Simulations</td>
<td>130</td>
</tr>
<tr>
<td>C.2.1</td>
<td>Generating Model Indicators</td>
<td>130</td>
</tr>
<tr>
<td>C.2.2</td>
<td>Generating $\lambda^{h(m+1)}$</td>
<td>131</td>
</tr>
<tr>
<td>C.2.3</td>
<td>Generating $\beta^h_s^{(m+1)}$</td>
<td>131</td>
</tr>
<tr>
<td>C.2.4</td>
<td>Generating $\mu^{(m+1)}_s$</td>
<td>132</td>
</tr>
<tr>
<td>C.2.5</td>
<td>Generating $\Sigma^{-1(m+1)}_s$</td>
<td>132</td>
</tr>
</tbody>
</table>
List of Tables

2.1 Estimation Results on Simulated Data .......................... 35
2.2 Parameter Estimates for MNL-ST ............................... 35
2.3 Sensitivity Analysis on Category Value Parameters ............ 36
2.4 Sensitivity Analysis on Price Parameters ....................... 37
2.5 Sensitivity Analysis on Price Parameters ....................... 38
2.6 Market Share/Average Price for the Yogurt Category .......... 39
2.7 Market Share/Average Price for the Ketchup Category ........ 40
2.8 NL and NL-ST Parameter Estimates and Fit Statistics .......... 41
2.9 HB-ST Parameter Estimates and Fit Statistics .................. 42
2.10 Regression of CV Parameters on Price Parameters ............ 43
2.11 Reference Price Elasticity for the Yogurt Category .......... 44
2.12 Reference Price Elasticity for the Ketchup Category ........ 45
2.13 Price Elasticity for the Yogurt Category ...................... 46
2.14 Price Elasticity for the Ketchup Category ...................... 47
3.1 Fit Statistics for the Yogurt Data .............................. 79
3.2 Fit Statistics for the Ketchup Data ............................. 80
3.3 Parameter Estimates for the Yogurt Data ....................... 81
3.4 Parameter Estimates for the Ketchup Data ....................... 82
3.5 FZENT: Yogurt Data ............................................ 83
3.6 Inclusion Probability and Market Share ......................... 84
3.7 Share Gains Across Brands: FZSET ............................. 85
3.8 Share Gains Across Brands: MNL

4.1 Parameter Estimates and Fit Statistics for HB-ST and -RD models

4.2 Parameter Estimates and Fit Statistics for HB-STRD model
List of Figures

3.1 Functional Shapes of Compensatory vs. Non-Compensatory Rule 87
3.2 Relationship Between Advertising and Price Sensitivity . . . . . 88
Acknowledgement

First of all, I thank God for helping me go through the ups and downs of completing my dissertation and giving me an opportunity to meet two wonderful advisors with high moral and academic standards.

Siddarth is the one with whom I have spent most of my time in developing ideas and translating them into my dissertation. Ongoing discussions with him, and his insightful suggestions, have helped me shape the core idea of my dissertation. Without his help I could have not finished my dissertation on time. Not only did he devote endless hours to reading the entire manuscript and helping me refine it, but he also demanded of me critical thinking and reasonable progress during the entire process. I could not have been more fortunate in finding a supervisor like Siddarth and I think of him as my big brother. Whoever has him as a supervisor is half-done in his/her dissertation.

Chuck Weinberg is the one whose part I would describe as that of an uncle in supervising my dissertation. His broad interests and knowledge in many areas in marketing have been a guiding light on the way I made progress at each stage. Seemingly, he was slow to understand the issues that I brought up in his presence (mainly because of my inarticulateness) but later on I was always amazed at his raising deeper issues in return, which I could not think of. Like an uncle, he has been always generous and encouraging (yet reasonably demanding). Again whoever has him as a supervisor is half-done in his/her dissertation.

Toward the end of the dissertation, their insightful reviews and suggestions I received helped to make significant refinements in this final version of the text. Whoever has both of them as supervisors is all-done in his/her dissertation.
Finally, the backbone to this entire effort is my family — my parents, for instilling in me the sense of perseverance needed to bring this dissertation to fruition, my wife, Sunghye, for appeasing our crying baby and putting up with the late night hours and my son, Jay, for being what I strive to work for.
Chapter 1

Introduction

1.1 Overview

With the advent of scanner panel data, discrete choice analysis in marketing has advanced our understanding of choice processes and consumer response to the marketing mix. The advantage of scanner panel data is that it represents actual behavior from a set of consumers, and offers the opportunity to observe the longitudinal choice behavior of panelists and a number of environmental variables in an unobtrusive way. This provides the researcher with an opportunity to either test theories or to use a scanner study to validate what has been shown in the laboratory. An important area of investigation in this research stream has been the development of methodologies to account for heterogeneity across households. One aspect of heterogeneity is the notion that consumers differ in their preference structure and in their response to the marketing mix. This has been incorporated by allowing the parameters of the brand choice model to vary across consumers. Several alternative methods have been proposed, notably, a latent class approach (Kamakura and Russell 1989), a fixed effects approach (Rossi and Allenby 1993) and a random effects approach (Chintagunta 1994, Jain, and Vilcassim 1991, Gonul and Srinivasan 1993, Rossi, McCulloch and Allenby 1996).

Another type of heterogeneity is derived from the notion that consideration or choice sets vary across consumers. Several approaches to model heterogeneity in consideration sets have been proposed in the literature (Andrews and Srinivasan 1995, Chiang,
Chapter 1. Introduction


While it is generally accepted that scanner panel data analysis should include parametric uncertainty and consideration set heterogeneity, much less attention has been paid to incorporating uncertainty about the model itself. That is, when faced with competing yet plausible structural assumptions, most often, analysts have used the data to identify a single best model and then proceeded as if the chosen model based on fit statistics were known to be true. The chosen model is then used to make inferences and predictions even though the rejected models are not necessarily wrong. In the statistics literature, this issue has been addressed by the Bayesian model averaging approach (Chatfield 1995, Draper 1995) that acknowledges that model uncertainty is likely to be more serious than other sources of uncertainty, e.g., parametric uncertainty, which have received far more attention from statisticians. The spirit of this approach is well represented in the following quote:

"The main way to avoid noticing after the fact that a set of modeling assumptions, different from those originally assumed, turned out to be correct is for one's model prospectively to have been sufficiently large to encompass the retrospective truth (Draper 1995, p. 55)."

In a broader sense, all kinds of heterogeneity corrections are reflections of the researcher's uncertainty about the true model structure. Therefore, as long as model complexity is manageable, the researcher seeks to build a bigger (more complex) model. Failure to do this may result in wrong inferences and misunderstanding of choice behavior.

In this thesis, we seek to examine the implications of heterogeneity in choice modeling,
Chapter 1. Introduction

i.e., parametric uncertainty, consideration set heterogeneity and structural uncertainty. More specifically, we investigate the relationship between reference price effects and parameter heterogeneity, the impact of advertising on price sensitivity through its influence on consideration set heterogeneity, and structural uncertainty about two competing reference price formulations: sticker shock (Winer 1986) and reference-dependent (Hardie, Johnson and Fader 1993).

Chapter 2 discusses the consequences of ignoring parametric uncertainty in investigating reference price effects. Several papers in the marketing literature have captured the impact of reference price on brand choice via the sticker shock formulation in which consumers are assumed to use the difference between the shelf price and reference price to evaluate an alternative. In this research, reference prices are brand-specific and typically imputed from the prices of the brands that a consumer is supposed to have “observed” on previous purchases in the category. However, since category marketing activity can also differentially affect the purchase timing of households, we argue that these measures of reference price may show certain systematic patterns and result in a spurious sticker shock effect. Specifically, we show that estimates of the sticker shock effect may be biased if households that are price sensitive in their brand choice decision are also more likely to respond to category marketing activity in their purchase timing decision. We present an analytical derivation to establish some general conditions under which the bias occurs and conduct a simulation experiment that confirms our specific hypotheses. We follow this up with an empirical application using scanner panel data from the yogurt and ketchup categories.

The results show that the 95% probability interval of the posterior distribution of the sticker shock coefficient in a Hierarchical Bayes model (which continuously accounts for price response heterogeneity across households) contains the value of zero, providing no evidence of the existence of a sticker shock effect. In contrast, the corresponding
coefficient from a standard model (which ignores this heterogeneity) is highly significant and supports the existence of a sticker shock effect. A posterior analysis on household parameters confirms the hypothesized positive relationship between responsiveness to category promotion activities and price sensitivity in the brand choice decision and is consistent with our explanation of the underlying cause of the bias in the standard model.

Chapter 3 presents a new "fuzzy" consideration set model that can be estimated with scanner panel. Unlike previous non-fuzzy (well-defined) set approaches which model uncertainty about the sets from which a brand is chosen, the proposed fuzzy set approach directly models the probability of a brand being included in the consideration set. The resulting inclusion probabilities for brands reflect a "fuzzy" consideration set, in the sense that a brand belongs to the consideration set only probabilistically, in contrast to the non-fuzzy consideration sets, in which the sets are well-defined but there is uncertainty about the sets from which a brand will be chosen.

As compared to an existing fuzzy set model (Bronnenberg and Vanhonacker 1996), in which variables have a compensatory influence on consideration, we model consideration as a non-compensatory process. We apply the approach to scanner panel data in two product categories and show that the proposed fuzzy set model outperforms several previous consideration set models.

We use our modeling approach to examine the role of the consideration set in moderating the effect of advertising on price sensitivity. This is to generalize the findings of Mitra and Lynch (1995), by using purchase data as opposed to experimental data, and also overcoming some of the limitations of their experimental study. Specifically, unlike their experimental study in which advertising could only increase consideration set size (since subjects only evaluated unfamiliar brands), we use a natural setting in which advertising can both increase and decrease consideration set size.

Empirical results show that consideration set size has no significant (positive) effect
Chapter 1. Introduction

on price sensitivity. The experimental findings of Mitra and Lynch (1995) do not seem to generalize to this particular data set. Further empirical test may be necessary to confirm their hypothesized relationship.

In marketing, a significant body of research has investigated the reference price construct and modeled its impact on consumer choice behavior via two major competing models, the sticker shock (Winer 1986), and reference-dependent formulations (Hardie, Johnson and Fader 1993). The sticker shock formulation captures reference price effects via a "sticker shock term" defined as the difference between brand-specific reference prices and the current prices. On the other hand, the reference-dependent formulation captures asymmetric responses to positive and negative price deviations from a category-specific reference price, which are hypothesized to result in loss aversion in the case of the current price exceeding the reference price (Tversky and Kahneman 1991). Therefore, these two models represent two different decision rules or choice processes underlying consumer brand choice, both of which are empirically supported as well as well-grounded in psychological theory: Adaptation-Level Theory (Helson 1964) and Prospect Theory (Kahneman and Tversky 1979).

Behaviorally, however, there is no reason to believe that every household follows the same decision rule. Some consumers may be well-described by the sticker shock model because they have sufficient cognitive capacity and motivation to recall the shelf-prices observed on the last purchase occasion. Other consumers may follow the reference-dependent model because they want to simplify the decision process, by using only one category-specific reference price to decrease the cognitive burden. One could also hypothesize a third segment of consumers that mixes the two decision rules probabilistically, i.e., uses one of the decision rules on different purchase occasions.

Previous research has not attempted to account for this structural uncertainty about
the way reference prices affect the choice process. Chapter 4 discusses the general problems of ignoring structural uncertainty and develops a Hierarchical Bayes model of reference price effects that accounts for heterogeneity in the way that consumers respond to reference prices in making brand choices.

Our empirical application of the proposed model to the ketchup category shows that consumers differ not only in their preference and response but also in their decision rules. On average, half the sample households show loss aversion while the remaining households do not respond to reference prices. The proposed model provides a richer description of consumer choice process than the base models that allow for only one model structure and ignore model uncertainty.
Chapter 2

Response Heterogeneity and Spurious Sticker Shock Effects

2.1 Overview

A significant body of recent research in marketing has investigated the reference price construct and modeled its impact on consumer brand choice behavior. Both aggregate (Raman and Bass 1988, Putler 1992) and disaggregate analysis (Winer 1986, Lattin and Bucklin 1989, Kalwani et al. 1990, Rajendran and Tellis 1994, Briesch et al. 1997) have been used to describe the nature of consumer response to reference price. In addition, a considerable body of work suggests an asymmetry in response to reference price with consumers treating gains differently from losses (Putler 1992, Hardie Johnson and Fader 1993). In their review paper, Kalyanaram and Winer (1995) interpret these, and other, research findings as strong support for the existence of reference price effects.

Previous disaggregate analysis has typically employed some variant of a logit choice model (Guadagni and Little 1983) and has used scanner panel data in model estimation. Since reference price is a latent construct and is difficult to measure through conventional data sources, it has been imputed from the price series in the data. In a majority of the models reported in the literature, the hypothesized impact of reference price is captured by including in the utility function of a logit choice model, a variable, commonly referred to as the “sticker shock” term (Winer 1986), that represents the difference between a brand’s reference price and its current shelf price. A positive coefficient for the associated parameter is interpreted as evidence for the “sticker shock” effect.
Chapter 2. Response Heterogeneity and Spurious Sticker Shock Effects

Virtually all previous sticker shock models have not accounted for the heterogeneity in price response among consumers (Bell and Bucklin 1998, and Briesch et al. 1997 are exceptions). The importance of capturing heterogeneity, and its implications for parameter estimates, has been well documented in the marketing literature (Chintagunta 1994, Jain and Vilcassim 1991, Kamakura and Russell 1989, Gonul and Srinivasan 1993, Rossi and Allenby 1993), and suggests that accounting for price response differences is an important prerequisite to obtaining accurate measures of the sticker shock effect.

In a study of employment decisions, Heckman (1981) also discusses the relationship between heterogeneity and state dependence. He argues that "individuals may differ in certain unmeasured variables that influence their probability of experiencing the event. If these variables are correlated over time, and are not properly controlled, previous experience may appear to be a determinant of future experience solely because it is a proxy for such temporally persistent unobservables" (Heckman 1981, p.91-92).

Bell and Lattin (1996) highlight a similar issue but, unlike the current work that examines the sticker shock effect, focus on the reference-dependent formulation and the role of unaccounted for heterogeneity in biasing estimates of loss aversion. They argue that price-sensitive consumers may predominantly observe losses (actual prices higher than a reference price) while price-insensitive consumers may predominantly observe gains (actual prices lower than a reference price). In this case, the best fit to the data is given by a model that allows for different response parameters for gains and losses. They show that, in a standard cross-sectional analysis, the parameter for losses (price-sensitive segment) tends to be greater in absolute magnitude than that for gains (price-insensitive segment) even in the absence of true loss aversion. In contrast to their work, we also

---
1In the reference-dependent formulation, the price of each choice alternative is compared to the same reference price, as opposed to the alternative-specific reference prices used in the sticker shock models. Consumers' responses to positive or negative deviations from this reference price are hypothesized to be asymmetric, resulting in loss aversion.
Chapter 2. Response Heterogeneity and Spurious Sticker Shock Effects

include an analytical derivation and a simulation study to describe the mechanism causing
the bias, and examine the role played by factors other than unaccounted for response
heterogeneity. Bell and Bucklin (1998) use a latent class approach to account for price
response heterogeneity in a sticker shock model of brand choice and purchase incidence.
They do not, however, discuss the bias issue but focus on the impact of reference prices
on the purchase incidence decision. Similarly, Briesch et al. (1997) also use the latent
class approach to compare alternative models of reference price.

While these studies have accounted for heterogeneity in parameters, the techniques
used have not been fully Bayesian. For example, the latent class approach used by Bell
and Bucklin (1998) and Briesch et al. (1997) approximates the continuous distribution
of heterogeneity by a discrete distribution. The Bayesian approach to fixed effects (Rossi
and Allenby 1993) used by Bell and Lattin (1996) requires *ad hoc* specification of prior
tightness. Since a thorough heterogeneity correction is crucial in examining “true” sticker
shock effects, we apply a more sophisticated Hierarchical Bayes approach (See Tanner
and Wong 1987, Gelfand and Smith 1990, Gelfand et al. 1990 for general discussion of
this method) to account for heterogeneity in preference and response in modeling the
brand choice decision.

Besides not accounting for response heterogeneity, a second, more subtle, character­
istic of previous research, is that the reference price measure is based on those trips on
which consumers *actually made a purchase in the category*. Such a measure, therefore,
directly includes the impact of pricing and promotion activity on the purchase timing
decision — an effect that has been extensively studied in the marketing literature (Gupta
1988, Jain and Vlccassim 1991, Bucklin, Gupta and Siddarth 1998). In this work, we
show that if the impact of promotional activity on a consumer’s category purchase deci­
sion depends on the underlying price response heterogeneity in the brand choice decision,
then the resulting reference price measure can cause biased estimates of the sticker shock
Chapter 2. Response Heterogeneity and Spurious Sticker Shock Effects

Effect. Specifically, we present an analytical derivation to show that this bias can occur if consumers who are more price-sensitive in the brand choice decision are also more likely to respond to promotional activity in the category in their purchase timing decision—a relationship for which there is some previous empirical evidence (Bucklin and Gupta 1992, Bucklin, Gupta and Siddarth 1998).

While unaccounted for price response heterogeneity alone may cause biased sticker shock estimates, results from a simulation study show that this requires very large variations in price responsiveness—unlikely to be observed in practice. In contrast, results from the simulation also demonstrate that, even moderate differences in brand choice price responsiveness, combined with systematic differences in purchase timing, are sufficient to produce biased estimates of the sticker shock effect.

This research seeks to provide the following contributions:

- Identify some general conditions under which unaccounted for price response heterogeneity can result in a spurious sticker shock effect.

- Describe the role that heterogeneity in purchase timing behavior plays in this process.

- Demonstrate, by way of an analytical derivation, a simulation study, and an analysis of scanner panel data, that properly accounting for price response heterogeneity can provide better measures of the sticker shock effect.

- Quantify the attenuation of the sticker shock effect as a result of heterogeneity correction.

This chapter is organized as follows. In Section 2.2 we present an analysis of a simple two segment setting to illustrate how the combination of unaccounted for price response heterogeneity and the impact of category promotion activity on purchase timing behavior
can result in a spurious sticker shock effect. Section 2.3 presents a simulation study that confirms the predictions from our theoretical analysis. In Section 2.4 we present an empirical application to scanner panel data from two product categories.

Results from our empirical analysis are consistent with our predictions. We find that the sticker shock coefficient is statistically significant in a standard model, which does not account for price response heterogeneity. On the other hand, the 95% probability interval\(^2\) of the posterior distribution of the sticker shock coefficient in a Hierarchical Bayes model contains the value of zero. Therefore, at least for the datasets used in this study, there is no evidence for the sticker shock effect. A posterior analysis of household parameters shows that, consistent with our explanation of the underlying cause of the spurious effects, households that are price sensitive in the choice decision are also more sensitive to category promotion activity in their decision to purchase in the category. Finally, we quantify the attenuation of the sticker shock effects via an elasticity analysis on estimated models. We close with our conclusions and directions for future research.

2.2 Theory

In this section, we discuss the conditions that can produce a spurious reference price effect. For ease of exposition, we consider the situation in which the market consists of two homogenous groups of consumers, differing in price sensitivity, but unresponsive to reference prices. We first specify the true model and follow with a description and analysis of the misspecified model.

\(^2\)A Bayesian probability interval for an unknown quantity of interest can be directly regarded as having high probability of containing the unknown quantity, in contrast to a frequentist's confidence interval, which is strictly interpreted only in the context of repeated sampling (Gelman et al. 1995).
2.2.1 Description of the True Model

We begin by specifying the deterministic component of household $h$'s utility for brand $i$ on occasion $t$ as,

$$U_{ht}^h = \alpha_i + \beta_c \text{PRICE}_{it},$$

(2.1)

where, $\alpha_i$ is a brand-specific constant, $\text{PRICE}_{it}$ is the price of brand $i$ on occasion $t$, and $\beta_c$ is the price response coefficient. To illustrate more clearly the logic of the underlying process, we omit other terms from the utility specification. The Multinomial Logit (MNL) model specifies the probability that household $h$ chooses brand $i$ on occasion $t$ as:

$$P_i^h(t) = \frac{\exp(U_{ht}^h)}{\sum_k \exp(U_{kt}^h)}.$$  

(2.2)

Including price response heterogeneity

Consider a situation in which consumers belong to one of two segments with either high, or low, price sensitivities, respectively (A straightforward extension to a situation for a more general distribution is provided in Appendix A). The utility function can then be modified as follows:

$$U_{ht}^h = \alpha_i + \beta_{hi} \text{PRICE}_{it} + \beta_d \text{SEG}^h \cdot \text{PRICE}_{it},$$

(2.3)

where $\text{SEG}^h$ is a dummy variable that takes the value 0 if household $h$ is in the high price-sensitive segment and 1 otherwise. The price parameter for the high price-sensitive segment is $\beta_{hi}(<0)$ and that for the low price-sensitive segment is $\beta_{hi} + \beta_d(<0)$, where $\beta_d > 0$. Equation (2.3) represents the true model that underlies the choice process and, in this formulation reference prices play no role in the household's choice decision.
2.2.2 Description of the Misspecified Model

We now take the position of an analyst who, incorrectly, ignores differences in price sensitivity across households, and omits the dummy variable, SEG\(^h\) from the utility function. Price responsiveness is captured by a single price parameter, \(\beta_c\), common to both segments. Also, an additional misspecification error is introduced by including in each brand's utility function a sticker shock term — the difference between its reference price and its current actual price. The resulting utility function, which is similar to that used in previous research, is

\[
U_{ht}^h = \alpha_i + \beta_c \text{PRICE}_{it} + \beta_{rp}(\text{RPRICE}_{it}^h - \text{PRICE}_{it}),
\]

(2.4)

which can be rewritten as,

\[
U_{ht}^h = \alpha_i + (\beta_c - \beta_{rp})\text{PRICE}_{it} + \beta_{rp}\text{RPRICE}_{it}^h,
\]

(2.5)

where \(\text{RPRICE}_{it}^h\) is a brand-specific reference price for household \(h\) on occasion \(t\). We expect \(\beta_c\) to lie in the range of the true price response parameters of each segment, i.e., \(\beta_{hi} < \beta_c < \beta_{hi} + \beta_d\). An estimate of \(\beta_{rp} > 0\) implies that consumers have reference prices and experience "sticker shock" in the brand choice decision.

Measuring reference price

Since reference price is an internal, latent, construct it has to be imputed from the prices paid (or observed) by the consumer on previous shopping occasions. The standard approach has been to operationalize reference price as a function of brand prices "observed" by the consumer on those trips on which a category purchase actually occurred. Kalyanaram and Little (1994) use an exponentially smoothed measure of past shelf prices encountered by a consumer and estimate the smoothing parameter to be 0.82, implying that the most recent price has the maximum influence on reference price. Briesch et al.
(1997) use a similar measure and report price history of about six periods in forming reference prices. Kalwani et al. (1990) summarize the impact of the previous five prices and find the most recent price to be weighted most heavily. Mayhew and Winer's (1992) exponential smoothing of past prices resulted in an optimal smoothing constant of one and led them to set reference price equal to that observed on the previous purchase occasion, i.e., $RPRICE^h_{it} = PRICE_{i(t-1)}$. A similar approach is also used by Bell and Bucklin (1998). Consistent with these findings, and earlier operationalizations of the reference price construct, we set the reference price of a brand equal to its price on the previous purchase occasion. While this definition has the added advantage of simplifying our exposition, we should point out that the underlying logic that we present carries over to other, more complex, measures of reference price. In the empirical analysis, we also re-estimate the model using a smoothed reference price measure and find the results to be unchanged.

### 2.2.3 Analysis of the Misspecified Model

In Appendix A, we present a formal analysis of the misspecified model and derive, in an OLS context, the conditions that can cause biased estimates of the sticker shock effect. Ideally, we could examine the bias issue in the setting of the logit choice model. However, the maximum likelihood estimates of model parameters are complicated nonlinear functions of data and do not provide much insight (See Yatchew and Griliches 1984 for a discussion of bias issues in the setting of the probit and logit models). Therefore, we choose to analyze bias in the OLS context. As we show in a later section, analytical results from a linear model are consistent with simulation results from the logit model.

The misspecified model involves the problem of omitted variables. Omission of a relevant variable from the model will bias the sticker shock term if What would cause the omitted variable to be also directly correlated to the reference price variable? Our
results show two general conditions that must be satisfied. First, consumers must differ in their price sensitivities, i.e., display heterogeneity in price response. Second, the omitted variable must be related significantly to both the utility (the left-hand side variable) and the sticker shock term. In the current instance, the omitted segment membership dummy variable is directly related to the brand utility derived by a household — since the utility depends on the price sensitivity. In this work we focus on the reference price measure, and its relationship to the purchase timing decision, as a possible cause of the bias. We now discuss the intuition underlying the results of our formal analysis.

**Role of Price Heterogeneity**

From Equation (2.5) it is clear that, for the segment of price sensitive households, any positive value of $\beta_{rp}$ increases its effective response to price by making $(\beta_c - \beta_{rp})$ more negative and, therefore, closer in magnitude to the true parameter value, $\beta_{hi}$. In other words, a positive estimate of the sticker shock parameter compensates for ignoring heterogeneity in price response. Moreover, this compensating mechanism also permits $\beta_c$ to be estimated closer to the true parameter value, $\beta_{hi} + \beta_d$, of the segment of less sensitive households. In summary, the presence of $\beta_{rp}$ makes the model more flexible in representing the underlying price heterogeneity — $\beta_c$ captures the response of the price-insensitive segment while $\beta_c$ and $\beta_{rp}$, together, capture the response of the price-sensitive segment.

**Role of Purchase Timing**

A significant body of previous research has investigated the impact of marketing activity on the purchase timing decision (Gupta 1988, Jain and Vilcassim 1991, Helsen and Schmittlein 1993). Bucklin and Gupta (1992) and Bucklin, Gupta and Siddarth (1998) provide some empirical evidence to show that price-sensitive segments have a greater
tendency to change purchase timing to coincide with promotional activity than price-insensitive segments. This empirical finding is of considerable significance to the current analysis.

Such a pattern of behavior implies that the price-sensitive segment is more likely to purchase in the category when it is more "attractive", i.e., when one or more brands is on promotion. The price-insensitive segment, on the other hand, is less likely to change its purchase timing due to promotional activity, resulting in a smaller proportion of its purchases coinciding with promotional activity. Therefore, the observed across-brand shelf prices will tend to be lower for the price-sensitive segment and higher for the insensitive segment. Since reference prices are set equal to observed prices on the previous purchase occasion, they will also display the same regularity. Equation (2.5) shows that the estimated coefficient for \( \beta_{rp} \) depends upon the measure of reference price used in the choice model. Also, as shown in Equation (A.8) (Appendix A), the bias in the reference price coefficient depends upon the relationship between the reference price measure and the underlying segment membership. Since the aggregate pattern of the reference prices across consumers, brands, and purchase occasions, varies systematically with the underlying segment membership (leading to a positive and significant \( \hat{\theta}_3 \) in Equation (A.9)), and the associated price heterogeneity, it may result in biased estimates of the sticker shock effect. It should be noted that if price response heterogeneity is present only in the brand choice decision (not in the purchase incidence decision), only one reference price\(^3\) (not the aggregate pattern of reference prices) would vary systematically with the segment membership, resulting in a smaller \( \hat{\theta}_3 \) and hence less serious bias.

Therefore, a tendency to change purchase timing plays a crucial role in biasing estimates.

---

\(^3\)Unobserved price sensitivity is revealed by the price actually paid for the one chosen brand which will be used as the reference price for the brand on the next purchase occasion. Other observed prices will be also used for reference prices, but will provide no information on the price sensitivity in the brand choice decision.
Chapter 2. Response Heterogeneity and Spurious Sticker Shock Effects

of the sticker shock effect.

**Extreme Price Heterogeneity**

If the two segments display extreme heterogeneity, i.e., a large $\beta_d$, in price responsiveness, it could result in a distinct pattern of prices in which the lower-priced brands are exclusively purchased by the price-sensitive segment and the higher-priced brands by the price-insensitive segment. This alternative mechanism could cause the reference price variable and the omitted variable to be correlated, even without differences in purchase timing, and result in biased estimates of the sticker shock parameter. In our simulation, we study the extent of heterogeneity required to produce the bias and find that the implied range of price elasticities is much larger than has been observed in practice (Tellis 1988a). In contrast, simulation results show that even moderate differences in price responsiveness, when combined with heterogeneity in purchase timing, can result in a biased sticker-shock parameter.

In summary, our analysis suggests that the combination of price response heterogeneity and the relationship between purchase timing and price sensitivity together make it more likely that the coefficient for the sticker shock term in the misspecified model is significant. Since reference prices are imputed from those occasions on which consumers actually purchase in the category, the purchase timing decision influences imputed reference prices which, in turn, biases estimates of the sticker shock effect. Also, while unaccounted for response heterogeneity in brand choice is a necessary condition for the existence of a spurious reference effect, that we can infer from Heckman (1981), it is not a sufficient one.
2.3 Simulation Study

Our theoretical analysis allows us to derive several predictions about the bias in the misspecified model. To confirm these predictions, we carried out a simulation study in which choices of hypothetical panelists are generated according to the true model, represented by Equation (2.3), after which the misspecified model is estimated on the same data.

2.3.1 Data Generation

We simulated category purchases, via a nested logit, and brand choice decisions, via a multinomial logit, for 400 hypothetical households of which 240 (60%) were assumed to be price-sensitive in the choice decision while the rest were price-insensitive. Each household was assumed to encounter each of the 60 different store environments, (i.e., make 60 shopping trips).

We generated 60 prices for each of two hypothetical brands of which the first brand (A) was assumed to be a national brand and the second (B) a non-promoted lower-priced brand. Regular prices for A were assumed to be uniformly distributed between $2.90 and $3.10, with a mean of $3, while the range for B was between $2.40 and $2.60, with a mean of $2.50. The combinations of prices for the two brands, therefore, represented 60 possible store environments that could potentially be encountered by the household on a shopping trip. The promoted price of Brand A was in the range of $2.30 to $2.50 (about 80% of its mean regular price) while Brand B was assumed never to be on promotion. The promotional frequency for Brand A was assumed to be about once in every three weeks, i.e., it was on promotion 20 out of 60 occasions. These levels are quite similar to those found in the yogurt and ketchup categories.

The following steps were used to generate the data. Each household was assumed to
encounter each of the 60, different, store environments, (i.e., make 60 shopping trips). For each store environment, we calculated the utility of each brand based on the following choice utility functions:

<table>
<thead>
<tr>
<th>Segment</th>
<th>Utility of Brand A</th>
<th>Utility of Brand B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Price Sensitive</td>
<td>$3 - \beta^h P_A$</td>
<td>$2 - \beta^h P_B$</td>
</tr>
<tr>
<td>Price Insensitive</td>
<td>$3 - \beta^l P_A$</td>
<td>$2 - \beta^l P_B$</td>
</tr>
</tbody>
</table>

In addition, the decision to purchase in the category was modeled using a nested logit in which the incidence utility was specified as a function of a constant and a category value term (CV) as follows:

$$V_s^* = a^* + \beta^s_{inc} \cdot CV,$$

where $\beta^s_{inc}$ is the response parameter for CV. CV for each trip is calculated as $\ln[\exp(U_A^s) + \exp(U_B^s)]$, and $s$ represents the segment to which the consumer belongs. The parameter $a^*$ was chosen such that the overall incidence probability for each household resulted in approximately 10 category purchases.

The probability of category purchase is given by:

$$P(inc) = \frac{\exp(V_s^*)}{1 + \exp(V_s^*)}.$$

A draw from the binomial vector $P(inc)$ was used to simulate an incidence decision. If the draw resulted in a predicted category purchase, then the brand choice decision was simulated using the multinomial logit choice probability. If the draw resulted in a non-purchase, we repeated the above steps for the next store environment. Thus, the resulting data set contained about 4000(10 x 400) “choice” observations.

*Price parameters in choice decision*

In a meta-analysis of 367 price elasticities, Tellis (1988a) reports elasticities in the range from +2 to -10 with a vast majority falling in the range -1 to -9. The choice
parameters used in the simulation reflect this range of elasticity differences. Thus, for the base case we choose the price parameters $\beta^h = -5$ and $\beta^l = -1$ which implies that price elasticities range from $-0.96$ to $-8.94$. In addition, we also analyzed several other parameter combinations outside this base range.

CV parameters in incidence decision

For each pair of choice parameters $(\beta^h, \beta^l)$, we generated data for two conditions. In the first, we explicitly allow the decision to purchase in the category to be linked to price sensitivity of the household, i.e., price-sensitive households are more likely to change their purchase timing decision based on price and promotion activity in the category. Consistent with results of earlier empirical analysis (Bucklin and Gupta 1992, Bucklin, Gupta and Siddarth 1998), the CV parameters were set at 0.7 and 0.2, for the high, and low price-sensitive segments, respectively. In the second condition, the category value response parameters were equal across segments ($\beta^h_{inc} = \beta^l_{inc} = 0.2$) reflecting price response heterogeneity without accompanying differences in the purchase timing responsiveness.

2.3.2 Predictions and Estimated Models

Our theoretical analysis allows us to make four predictions about parameters of models estimated on the simulated data. First, in the presence of unaccounted for heterogeneity, the parameter $\beta_{rp}$ will tend to be positive and significant. Second, when compared to the estimate obtained from a model which omits the sticker shock term (a "plain" MNL model), the common price parameter, $\beta_c$, which tends to compensate for the price-insensitive segment, will have a smaller absolute value. Third, a latent class model, which accounts for price response heterogeneity but (incorrectly) includes a sticker shock term, will not result in a spuriously significant sticker shock coefficient. Fourth, the misspecified model which does not account for price response heterogeneity will show no
bias in the sticker shock parameter if the underlying heterogeneity in purchase timing is the same across the segments. To test these predictions, we estimate three models on the simulated data, the “plain” MNL model, the corresponding sticker shock model (MNL-ST) and a multi-segment latent class model, which (mistakenly) includes a sticker shock term (SEG-ST).

2.3.3 Simulation Results

Base case results

*CV parameters different across segments*

Maximum likelihood estimation results for the base case ($\beta^h = -5$ and $\beta^t = -1$) are displayed in Table 2.1. The price coefficient ($-2.89$) for the MNL model lies near the midpoint of the true values ($-1$ and $-5$). The coefficient for the sticker shock term in the MNL-ST model is positive and significant even though the true model does not include a reference price term in the utility function. Additionally, in line with our predictions, the price coefficient for the MNL-ST model has a smaller absolute value than its counterpart in the MNL model ($2.61 < 2.89$). These results confirm that introducing the sticker shock term increases model flexibility and permits the common coefficient, $\beta_c$, to be estimated closer to the true price coefficient of the insensitive segment. Finally, the results from the latent class model, SEG-ST, show that accounting for price response heterogeneity via separate price coefficients for each segment causes the spurious sticker shock effect to disappear.\(^4\)*

*CV parameters equal across segments*

Model results for the MNL-ST fitted to the base case data are presented in Table

\(^4\)We also generated data for the case in which the true model included sticker shock terms with known parameters (1.0 for the high sensitive segment and 0.5 for the low sensitive segment, respectively) in both segments and found that the latent class model was successful at recovering the true sticker shock parameters.
2.2. In contrast to the results from the previous simulation, we find that the coefficient for the sticker shock term is insignificant. These results show that unaccounted for heterogeneity in price response is a necessary, but not sufficient, condition to obtain a spuriously significant sticker shock effect.

Robustness check

We estimated the MNL-ST model on fifty additional data sets generated for each of several different combinations of the CV parameters, $\beta_{inc}^h$ and $\beta_{inc}^l$. Results, presented in Table 2.3, show that the mean bias increases as the differences in the CV response parameters across the two segments increases, and is statistically insignificant when the value of the difference term is less than 0.4. These results confirm that differences in purchase timing responsiveness play an important role in determining the extent of bias.

Price Parameters Different From the Base Case

CV parameters different across segments

Table 2.4 reports the mean value of the sticker shock parameter, $\beta_{rp}$, and the associated t-statistic obtained from estimating the MNL-ST model on fifty different data sets generated for each of seven parameter combinations. The results indicate that term $\beta_{rp}$ is insignificant when differences in heterogeneity are small but becomes significant as price heterogeneity increases.

CV parameters equal across segments

Results for this case are reported in Table 2.5. Only very extreme variations in the price parameters result in a spurious sticker shock effect. The implied range of elasticities is much higher than that reported in previous research (Tellis 1988). This contrasts with

---

5If $\hat{\beta}_{rp}^t$ is the estimate obtained from a single data set $t$, and $\bar{\beta}_{rp}$ represents the mean across the 50 simulations then we can write a standard error as $S.E.(\bar{\beta}_{rp}) = \sqrt{\frac{\sum_{t=1}^{50} \text{Var}(\hat{\beta}_{rp}^t)}{50}}$
our earlier findings in which even moderate differences in price responsiveness, when combined with purchase timing heterogeneity, resulted in a spurious sticker-shock coefficient.

2.4 Empirical Application

To test our theory further, we conduct an empirical test using scanner panel data from the yogurt and ketchup categories. We develop and estimate a Hierarchical Bayes model of incidence and sticker shock choice (labeled as HB-ST) that accounts for heterogeneity in all parameters across consumers. Since the Hierarchical Bayes model eliminates the source of bias, we expect the magnitude of the coefficient for the sticker shock term to be attenuated and to represent the true impact of reference price on brand choice.

For comparison purposes, we estimate a standard nested logit model of incidence and choice (Bucklin and Gupta 1992), labeled NL, as well as a nested logit model that includes a sticker shock term in the choice model, labeled as NL-ST, that does not account for price response heterogeneity. We expect the sticker shock coefficient from the NL-ST model to be positive and statistically significant.

2.4.1 Data

The data for this study are drawn from A.C. Nielsen scanner panel records in the yogurt and ketchup categories, for households in Sioux Falls, South Dakota, for the period 1986-1988. The last 51 weeks of the data were used for model calibration and the preceding 61 weeks were used for initializing model variables. Households qualified for inclusion in the sample if they made at least one grocery purchase every four weeks over the entire study period and made at least one product purchase both in initialization and calibration periods. Out of the qualified households, 400 were randomly selected from each product category for model calibration (42954 shopping trips and 3852 purchases in the yogurt
category, 44631 shopping trips and 2476 purchases in the ketchup category).

The top seven selling brands are included in the study. Together, these accounted for 84 (85) percent of category sales in dollars, and 89 (87) percent in units for the yogurt and ketchup categories, respectively. These brands, their market shares by units, and average prices per oz are given in Table 2.6 and Table 2.7.

2.4.2 Models Estimated

Nested Logit (NL) Model

This model assumes that, conditional on purchase incidence, consumers choose a brand without relying on internal reference price and that they share the same response parameters. The NL model specifies the probability of brand choice for household $h$ on occasion $t$ as:

$$ P_t^h(i|inc) = \frac{\exp(U_{it}^h)}{\sum_k \exp(U_{kt}^h)}. $$

(2.6)

$U_{it}^h$ denotes the deterministic component of utility for each alternative $i$ and is specified as a function of both household-specific and marketing mix variables as follows:

$$ U_{it}^h = \beta_0 + \beta_1 \text{BLOY}_i^h + \beta_2 \text{LBP}_{it}^h + \beta_3 \text{PRICE}_{it} + \beta_4 \text{PROMO}_{it}, $$

(2.7)

where,

- $\text{BLOY}_i^h$ = within-household market share of brand $i$ in the initialization period,
- $\text{LBP}_{it}^h$ = 1 if $i$ was the brand purchased on the previous occasion; 0 otherwise,
- $\text{PRICE}_{it}$ = actual shelf price of brand $i$ at time $t$, and
- $\text{PROMO}_{it}$ = 1 if brand $i$ is on promotion at time $t$; 0 otherwise.

The above formulation accommodates cross-sectional heterogeneity in brand preference through the static brand loyalty measure and time-varying heterogeneity through the last purchase variable, $\text{LBP}_{it}^h$. 
Chapter 2. Response Heterogeneity and Spurious Sticker Shock Effects

A consumer's category purchase decision is modeled with a binary nested logit. Specifically, the probability of purchase incidence for household \( h \) on a store visit at time \( t \) is given by (e.g., Bucklin and Gupta 1992):

\[
P_t^h(inc) = \frac{\exp(V_t^h)}{1 + \exp(V_t^h)}. \tag{2.8}
\]

\( V_t^h \) denotes the deterministic component of utility for purchase incidence and is specified as follows:

\[
V_t^h = \gamma_0 + \gamma_1 CR_t^h + \gamma_2 INV_t^h + \gamma_3 CV_t^h. \tag{2.9}
\]

where,

\[
CR_t^h = \text{average weekly consumption rate for household } h,
\]

\[
INV_t^h = \text{inventory estimate for household } h \text{ at time } t,
\]

\[
CV_t^h = \text{category value at time } t \text{ for household } h.
\]

Consumption rate is computed as the total amount of product purchased by a household \( h \) in the initialization period divided by the number of weeks in the period. It remains constant through the calibration period.

The inventory variable, \( INV_t^h \), is designed to capture time-varying heterogeneity in incidence probabilities. To construct the inventory variable, we assume that households draw down their supply at their rates of consumption, \( CR_t^h \). We initialize our inventory measure at zero at the start of the initialization period. We also mean-center \( INV_t^h \) by subtracting each household's average level of inventory during the calibration period. This makes the measure purely longitudinal so that \( INV_t^h \) becomes a measure of relative inventory within a household (Bucklin and Gupta 1992).

In the nested logit incidence model, the attractiveness of the product category (due to the price and promotion of the various brands) is represented by category value, \( CV_t^h \) (the log of the denominator of the brand choice model). The coefficient for category value
should be positive (reflecting the positive influence of, say, a price promotion on the incidence decision) and lie between zero and one so that it is consistent with stochastic utility maximization theory (Ben-Akiva and Lerman 1985).

**Nested Logit Sticker Shock (NL-ST) Model**

The NL-ST model calls for augmenting the NL brand choice utility function with the sticker shock term that reflects the difference between a brand's reference price and its shelf price. Following Winer (1986) and Lattin and Bucklin (1989), the sticker shock utility is given by:

\[
U_{it}^h = \beta_0 + \beta_1 \text{BLOY}_{i}^h + \beta_2 \text{LBP}_{it}^h + \beta_3 \text{PRICE}_{it} + \beta_4 (\text{RPRICE}_{it}^h - \text{PRICE}_{it}) + \beta_5 \text{PROMO}_{it},
\]

(2.10)

where \(\text{RPRICE}_{it}^h\) is a brand-specific reference price. The reference price of a brand is assumed to be its price on the last occasion on which the household made a purchase in the category. The brand choice and purchase incidence probabilities are the same as in Equation (2.6) and Equation (2.8), respectively.

**Hierarchical Bayes Sticker Shock (HB-ST) Model**

The heterogeneity in choice and incidence parameters is modeled via the HB-ST model. The model has three stages. In the first stage, the brand choice and purchase incidence models are written with household-specific parameters. In the second stage, the prior (population) distribution over the household-specific parameter vectors, \(\theta^h = (\beta^h, \gamma^h)\), is specified as a multivariate normal, i.e., \(\theta^h \sim MVN(\mu, \Sigma)\). The covariance matrix of household parameter vectors, \(\Sigma\), captures the unobserved source of heterogeneity across households. Since the fully Bayesian approach requires priors over unknown quantities in the model, the third stage of the model specifies hyper-priors over the parameters of
Chapter 2. Response Heterogeneity and Spurious Sticker Shock Effects

the population distribution, $\mu$ and $\Sigma$. We assume a multivariate normal hyper-prior, $MVN(\eta, C)$ over $\mu$ and a Wishart prior, $W[(\rho R)^{-1}, \rho]$ over $\Sigma$ following the standard approach suggested in the statistics literature (Gelfand and Smith 1990, Gelfand et al. 1990). Details on each prior are provided in Appendix B.

This hierarchical set up involves evaluating high-order multivariate normal integrals required in formulating the posterior density, making it practically impossible to use conventional maximum likelihood estimation techniques. We circumvent this problem by using sampling-based Bayesian approaches that build upon recent advances in Markov Chain Monte Carlo (MCMC) methods (Gelfand and Smith 1990, Gelfand et al. 1990). Using the sampling-based approach permits one to account properly for uncertainty in inferences and to make exact (as opposed to asymptotic) inference possible with any desired precision (since the number of draws is under an analyst's control).

2.4.3 Estimation

**NL & NL-ST Models**

The NL and NL-ST models have the following log-likelihood function:

$$LL(\theta) = \sum_{h=1}^{H} \left[ \sum_{t=1}^{T_h} \ln \left[ P_t^h(\text{inc})^{\delta_t^h} (1 - P_t^h(\text{inc}))^{1 - \delta_t^h} \prod_i \left[ P_t^h(i|\text{inc})\right]^{\delta_{it}^h} \right] \right]$$  \hspace{1cm} (2.11)

where $\delta_{it}^h = 1$ if household $h$ bought brand $i$ on a store visit at time $t$, and zero otherwise, and $\delta_t^h = \sum_i \delta_{it}^h$. $\theta$ is the choice and incidence parameter vector which is common across households. Maximization of Equation (2.11) simultaneously yields the choice and incidence parameters.

**HB-ST Model**

Sampling from the joint posterior distribution is achieved by sampling from the full conditional distributions, known as substitution sampling (Details on the full conditional
Chapter 2. Response Heterogeneity and Spurious Sticker Shock Effects

distributions for model parameters are provided in Appendix B). If all the full conditional distributions are known (i.e., known conjugate distributions such as a normal and a Wishart), then substitution sampling reduces to a procedure known as Gibbs Sampling (Geman and Geman 1984, Gelfand and Smith 1990). If the full conditional distributions are not completely known (i.e., known only up to a normalizing constant), then a Metropolis-Hastings step is used (Tierney 1994, Chib and Greenberg 1995). In the context of the HB-ST model developed in the previous section, the \((m + 1)\)th step of the substitution sampling involves generating the following draws:

(a) Generate \(\theta^h\) draws from \(p(\theta^{h(m+1)}|\Sigma^{(m)}, \mu^{(m)})\) for \(h = 1\) to \(H\), using a Metropolis-Hastings algorithm.

(b) Generate a \(\mu\) draw from \(p(\mu^{(m+1)}|\{\theta^{h(m)}\}, \Sigma^{(m)}, \eta, C)\), using a Gibbs sampler.

(c) Generate a \(\Sigma^{-1}\) draw from \(p(\Sigma^{-1(m+1)}|\{\theta^{h(m)}\}, \mu^{(m)}, \rho, R)\), using a Gibbs sampler.

This sequence of draws generates a Markov chain whose stationary distribution is the joint posterior distribution of all unknown parameters. The initial draws (1st to \(m\)th) from the chain reflect a transient period in which the chain has not converged to the equilibrium distribution and is therefore discarded. A sample of draws obtained after convergence is used to make posterior inferences about model parameters. It should be noted that this sampling-based procedure generates the entire posterior distribution rather than just a point estimate. Given that we expect substantial uncertainty in inferring household parameters with only a small number of observations, it is critical that we should also measure properly the uncertainty of estimates.

In our application, the substitution sampler was run for 15000 iterations and convergence was ensured by monitoring the time-series of the draws. We chose a burn-in length of 10000 iterations (after which the time-series plot stabilized.) and retained every fifth
iteration of the remaining 5000 iterations to reduce the serial correlation of the sampled draws, i.e., resort to “thinning the chain” (Geyer 1992, Raftery and Lewis 1995). Therefore, 1000 draws from the posterior distribution of each parameter were used to make inferences.

2.5 Results and Discussion

In this section, we compare the results of the standard (NL and NL-ST) and HB-ST models. Traditional measures of goodness-of-fit such as the Bayesian Information Criteria (BIC) and Akaike Information Criteria (AIC), which penalize the likelihood for the number of estimated parameters, are not suitable for assessing and comparing Hierarchical Bayes models against non-Bayesian models because these measures are motivated by asymptotic arguments and are, therefore, inappropriate when the number of model parameters varies with the sample size (Manchanda et al. 1997, Carlin and Louis 1996), as in the HB-ST model. Therefore, we use log-likelihood and in-sample hit-rates\(^6\) to compare the models.

2.5.1 NL and NL-ST models

Fit statistics and parameter estimates are presented in Table 2.8. The NL-ST model provides a better fit than the NL model in both categories. The sticker shock coefficient is positive and significant in both categories. Also, as per our prediction, in both categories, the common price parameter, \(\beta_c\), in the NL-ST model is larger than the corresponding value in the NL model.

\(^6\)Each household is assumed to pick the brand with the highest predicted choice probability. The percentage of correct choice predictions gives the hit-rate (Kalwani, Meyer and Morrison, 1994).
2.5.2 HB-ST Model

Estimates of the population parameters ($\mu$), and fit statistics, are shown in Table 2.9. In both categories, the HB-ST model outperforms the baseline models in terms of both log-likelihood and hit-rates. As is usual for Bayesian inference, we summarize the posterior distribution of the parameters by the posterior mean and 95% probability interval for each parameter. In both categories, this interval for the sticker shock coefficient contains the value zero. Thus, properly accounting for price response heterogeneity mutes the impact of reference price and, in our application, the sticker shock effect disappears completely. For these data sets therefore, the results do not support the existence of a sticker shock effect. In addition, the magnitude of another state-dependent variable, LBP, is substantially smaller in both categories supporting the notion that state dependence is attenuated by accounting for heterogeneity.\(^7\)

2.5.3 Posterior Analysis

As discussed previously, a spurious reference price effect requires households that are price-sensitive in the choice decision also to be more responsive to price and promotion in the category purchase decision. To provide support to this argument, household CV parameters were regressed on the price parameters.\(^8\) The regression results, presented in Table 2.10, show that, in both categories, greater responsiveness to category value is associated with higher price sensitivity in the choice decision, providing strong empirical

\(^7\)We re-estimated the NL-ST model using an exponentially smoothed reference price measure, $\text{RP}_{it} = \text{RP}_{i(t-1)}(1 - \alpha) + \text{RP}_{i(t-1)}\alpha$, similar to that used by Briesch et. al. (1997). This measure provided a very slight improvement (about one likelihood point) in fit over the current measure, $\text{RP}_{it} = P_{i(t-1)}$ and $\alpha$ was estimated to be 0.72 (0.56) for the yogurt (ketchup) category. We then re-estimated the HB-ST model using this definition of reference prices, and found our results to be unchanged, i.e., the 95% probability interval for the sticker shock parameter still contained the value zero.

\(^8\)The substitution sampling yields a sample of 1000 draws for all model parameters. We then obtain point estimates of household parameters ($\theta^h$) by averaging over this distribution. Specifically, we regress $\gamma^h_3$ on $\beta^h_3$. 
support for our theory.

### 2.5.4 Attenuation of Reference Price Effects

In both categories, we find that the sticker shock effect from the standard (homogeneous) model is overestimated. We are interested in quantifying the extent of attenuation due to heterogeneity correction. One straightforward way to do this is by comparing the elasticity of reference price of the standard model to that of the HB-ST model.

For the HB-ST model, we base our inference about elasticities on a sample of 1000 draws from the posterior distribution of population parameters ($\mu$). For each draw, we calculate the reference price elasticities across purchase occasions and average them across brands. Then we repeat the procedure for 1000 draws and average the brand-level elasticities across draws. We present the elasticity comparisons between the models for both categories in Table 2.11 and Table 2.12.

Results show that attenuation of the effect size is quite significant with average percent decreases being 63% and 61% for the yogurt and ketchup categories, respectively (It should be noted that in fact, the value zero is included in 95% probability intervals of all elasticity estimates of the HB-ST model, indicating an insignificant sticker shock effect). In contrast, price elasticities have increased in the absolute magnitude when heterogeneity is accounted for, as presented in Table 2.13 and Table 2.14. Therefore, accounting for heterogeneity appears to sharpen the effect of price while attenuating the sticker shock effect.

### 2.6 Conclusion

The reference price construct, and its role in the brand choice decision, has been extensively investigated by researchers in marketing using secondary data from scanner panels.
Chapter 2. *Response Heterogeneity and Spurious Sticker Shock Effects*

Consumer heterogeneity also has a rich modeling tradition in the marketing literature and several papers have highlighted the importance of accounting for it in order to obtain unbiased parameter estimates from choice models. However, most previously reported models of reference price effects have not adequately accounted for heterogeneity.

The current work studies the impact of unaccounted for price response heterogeneity on estimates of the sticker shock effect. The analysis reveals two conditions under which estimates of the sticker shock effect are biased. First, the estimated model must ignore true differences in price sensitivities among consumers. Second, one, or more, of the other explanatory variables used in the analysis should systematically vary with the underlying price responsiveness. This research highlights the role that heterogeneity in purchase timing can play in this process. Based on this analysis, several predictions about the nature of parameter bias are made and confirmed via a simulation study. Simulation results also show that, while the existence of extreme (unaccounted for) price response heterogeneity can, in itself, cause biased estimates of the sticker shock effect, the implied range of price response variation is much larger than observed in practice.

A Hierarchical Bayes version of the nested logit model is proposed and estimated on scanner panel data from two product categories. In contrast to a nested logit model with a sticker shock term, which does not account for price response heterogeneity, the 95% confidence interval of the sticker shock parameter in the HB-ST model includes the value zero. For both these data sets, therefore, the results do not support the existence of a brand-specific reference price. Household-level category value and price response parameters obtained from the posterior distribution, show that greater responsiveness to category value is associated with higher price sensitivity in the choice decision. Overall, these results provide strong support for our theory of the determinants of sticker-shock bias.
Our results complement those of Bell and Lattin (1996) who focus on the reference-dependent formulation of reference price. It is interesting to note that their explanation for bias is consistent with our analysis. Thus, in their work, gains and losses mimic the pattern of price heterogeneity of the underlying segments and play the same role as the sticker shock term does in our model. Our empirical work in two product categories (yogurt and ketchup) shows that accounting for price response heterogeneity results in the sticker shock effect being insignificant. Their results also show that the loss aversion is significantly attenuated, though it does not completely disappear.

Posterior analysis on reference price elasticities show that attenuation of the effect size after accounting for heterogeneity is quite substantial with average percent decreases being 63% and 61% for the yogurt and ketchup categories, respectively. However, given large uncertainty bounds around elasticity estimates (i.e., 0 is included in 95% probability intervals of the estimates), real effect sizes are considered to be negligible once price response heterogeneity is accounted for. In these particular categories, our results, therefore, do not support the existence of a brand-specific reference price as operationalized according to the standard approach in previous research.

Recent work in the marketing literature has used reference price effects to derive optimal promotion and pricing policies (Koppalle, Rao and Assuncao 1996, Greenleaf 1995). Since accurate estimates of the size and prevalence of reference price effects are critical to the development of these strategies, our research highlights the importance of accounting for consumer heterogeneity in obtaining such estimates.

2.7 Limitations and Future Research

Future research could take several directions. Laboratory and survey work could be used to uncover the mechanisms that consumers actually use to form reference prices
in different product categories. Besides providing a useful framework to understand the construct, this research may identify better heuristics that could be used to construct reference price measures in studies using secondary data sources. Another fruitful area for research would be to compare the results obtained by different methods of accounting for consumer heterogeneity. In fact, when we employed a latent class approach (Kamakura, Kim and Lee 1996), six and three segments, respectively, emerged in the yogurt and ketchup categories. Our results regarding the sticker shock effect was the same in the yogurt category but one of the segments (represents 29% of the sample) in the ketchup category showed a significant sticker shock effect. This implies that residual price response heterogeneity within a segment could potentially result in a spurious sticker shock effect.

The objective of this research was to describe the mechanism by which price response heterogeneity could bias estimates of the sticker shock effects of reference price via a more complete treatment of heterogeneity than carried out in the previous studies. Our work points out some of the limitations of previous studies and shows that considerable additional analysis, perhaps over multiple categories, will be needed to describe accurately the formation and role of reference prices in brand choice.
Table 2.1: Estimation Results on Simulated Data
Category Value Parameters Different Across segments
$\beta^h_{inc} = 0.7, \beta^l_{inc} = 0.2$

<table>
<thead>
<tr>
<th>Variable</th>
<th>MNL</th>
<th>MNL-ST</th>
<th>SEG-ST</th>
</tr>
</thead>
<tbody>
<tr>
<td>PRICE1</td>
<td>-2.89</td>
<td>-2.61</td>
<td>-5.18</td>
</tr>
<tr>
<td></td>
<td>(-24.29)</td>
<td>(-16.25)</td>
<td>(-12.33)</td>
</tr>
<tr>
<td>RPRICE1-PRICE</td>
<td>0.29</td>
<td>0.04</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.53)</td>
<td>(0.17)</td>
<td></td>
</tr>
<tr>
<td>PRICE2</td>
<td></td>
<td>-1.07</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-3.62)</td>
<td></td>
</tr>
<tr>
<td>RPRICE2-PRICE</td>
<td>0.17</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.88)</td>
<td></td>
</tr>
</tbody>
</table>

Asymptotic $t$-statistics in parentheses.

Table 2.2: Parameter Estimates for MNL-ST
Category Value Parameters Equal Across Segments
$\beta^h_{inc} = 0.2, \beta^l_{inc} = 0.2$

<table>
<thead>
<tr>
<th>Variable</th>
<th>Parameter Estimate</th>
<th>t-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>PRICE</td>
<td>-2.95</td>
<td>-17.47</td>
</tr>
<tr>
<td>RPRICE-PRICE</td>
<td>0.06</td>
<td>0.48</td>
</tr>
</tbody>
</table>
Table 2.3: Sensitivity Analysis on Category Value Parameters

Base Case
\( \beta^h = -5, \beta^l = -1 \)

<table>
<thead>
<tr>
<th>( \beta^h_{inc} )</th>
<th>( \beta^l_{inc} )</th>
<th>( \beta^h_{inc} - \beta^l_{inc} )</th>
<th>Mean ( \beta_{rp} )</th>
<th>t-value for Mean ( \beta_{rp} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.7</td>
<td>0.2</td>
<td>0.5</td>
<td>0.29</td>
<td>2.53</td>
</tr>
<tr>
<td>0.6</td>
<td>0.2</td>
<td>0.4</td>
<td>0.25</td>
<td>2.15</td>
</tr>
<tr>
<td>0.5</td>
<td>0.2</td>
<td>0.3</td>
<td>0.20</td>
<td>1.75</td>
</tr>
<tr>
<td>0.4</td>
<td>0.2</td>
<td>0.2</td>
<td>0.15</td>
<td>1.30</td>
</tr>
<tr>
<td>0.3</td>
<td>0.2</td>
<td>0.1</td>
<td>0.11</td>
<td>0.93</td>
</tr>
<tr>
<td>0.2</td>
<td>0.2</td>
<td>0.0</td>
<td>0.06</td>
<td>0.48</td>
</tr>
</tbody>
</table>
Table 2.4: Sensitivity Analysis on Price Parameters
Category Value Parameters Different Across Segments
\( \beta_{inc}^h = 0.7, \beta_{inc}^l = 0.2 \)

<table>
<thead>
<tr>
<th>Price Parameter for Segment 1</th>
<th>Price Parameter for Segment 2</th>
<th>Implied Elasticity Range</th>
<th>Mean ( \beta_{rp} )</th>
<th>t-value for Mean ( \beta_{rp} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1</td>
<td>0</td>
<td>-1.67 to 0</td>
<td>0.039</td>
<td>0.33</td>
</tr>
<tr>
<td>-1</td>
<td>-2</td>
<td>-2.94 to -0.96</td>
<td>0.037</td>
<td>0.34</td>
</tr>
<tr>
<td>-1</td>
<td>-3</td>
<td>-4.26 to -0.96</td>
<td>0.091</td>
<td>0.82</td>
</tr>
<tr>
<td>-1</td>
<td>-4</td>
<td>-6.50 to -0.96</td>
<td>0.194</td>
<td>1.72</td>
</tr>
<tr>
<td>-1</td>
<td>-5</td>
<td>-8.94 to -0.96</td>
<td>0.287</td>
<td>2.53</td>
</tr>
<tr>
<td>-1</td>
<td>-10</td>
<td>-20.68 to -0.96</td>
<td>0.501</td>
<td>4.14</td>
</tr>
<tr>
<td>-1</td>
<td>-20</td>
<td>-40.95 to -0.96</td>
<td>0.552</td>
<td>4.30</td>
</tr>
</tbody>
</table>
Table 2.5: Sensitivity Analysis on Price Parameters
Category Value Parameters Equal Across Segments
$\beta_{inc}^h = 0.2, \beta_{inc}^l = 0.2$

<table>
<thead>
<tr>
<th>Price Parameter for Segment 1</th>
<th>Price Parameter for Segment 2</th>
<th>Implied Elasticity Range</th>
<th>Mean $\beta_{rp}$</th>
<th>t-value for Mean $\beta_{rp}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1</td>
<td>0</td>
<td>-1.67 to 0</td>
<td>0.013</td>
<td>0.11</td>
</tr>
<tr>
<td>-1</td>
<td>-2</td>
<td>-2.94 to -0.96</td>
<td>-0.010</td>
<td>-0.09</td>
</tr>
<tr>
<td>-1</td>
<td>-3</td>
<td>-4.26 to -0.96</td>
<td>-0.014</td>
<td>-0.12</td>
</tr>
<tr>
<td>-1</td>
<td>-4</td>
<td>-6.50 to -0.96</td>
<td>0.004</td>
<td>0.04</td>
</tr>
<tr>
<td>-1</td>
<td>-5</td>
<td>-8.94 to -0.96</td>
<td>0.056</td>
<td>0.48</td>
</tr>
<tr>
<td>-1</td>
<td>-10</td>
<td>-20.68 to -0.96</td>
<td>0.170</td>
<td>1.40</td>
</tr>
<tr>
<td>-1</td>
<td>-20</td>
<td>-40.95 to -0.96</td>
<td>0.280</td>
<td>2.20</td>
</tr>
</tbody>
</table>
Table 2.6: Market Share/Average Price for the Yogurt Category

<table>
<thead>
<tr>
<th>Brand</th>
<th>Market Share</th>
<th>Average Price per Oz.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yoplait</td>
<td>0.19</td>
<td>10.1 cents</td>
</tr>
<tr>
<td>Weight Watcher</td>
<td>0.12</td>
<td>7.7 cents</td>
</tr>
<tr>
<td>Dannon</td>
<td>0.08</td>
<td>8.5 cents</td>
</tr>
<tr>
<td>Nordica</td>
<td>0.29</td>
<td>6.7 cents</td>
</tr>
<tr>
<td>QC</td>
<td>0.03</td>
<td>5.2 cents</td>
</tr>
<tr>
<td>W.B.B</td>
<td>0.17</td>
<td>5.4 cents</td>
</tr>
<tr>
<td>Private Label</td>
<td>0.11</td>
<td>4.8 cents</td>
</tr>
</tbody>
</table>
### Table 2.7: Market Share/Average Price for the Ketchup Category

<table>
<thead>
<tr>
<th>Brand</th>
<th>Market Share</th>
<th>Average Price per Oz.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hunt's 32 Oz.</td>
<td>0.14</td>
<td>3.45 cents</td>
</tr>
<tr>
<td>Del Monte 32 Oz.</td>
<td>0.07</td>
<td>3.46 cents</td>
</tr>
<tr>
<td>Heinz 28 Oz.</td>
<td>0.23</td>
<td>4.52 cents</td>
</tr>
<tr>
<td>Heinz 32 Oz.</td>
<td>0.40</td>
<td>3.51 cents</td>
</tr>
<tr>
<td>Heinz 40 Oz.</td>
<td>0.06</td>
<td>4.83 cents</td>
</tr>
<tr>
<td>Heinz 64 Oz.</td>
<td>0.04</td>
<td>4.60 cents</td>
</tr>
<tr>
<td>Private Label</td>
<td>0.06</td>
<td>2.76 cents</td>
</tr>
</tbody>
</table>
Table 2.8: NL and NL-ST Parameter Estimates and Fit Statistics

<table>
<thead>
<tr>
<th>Variable</th>
<th>Yogurt</th>
<th>Ketchup</th>
<th>Yogurt</th>
<th>Ketchup</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NL</td>
<td>NL-ST</td>
<td>NL</td>
<td>NL-ST</td>
</tr>
<tr>
<td>Brand Choice</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BLOY</td>
<td>1.701</td>
<td>1.694</td>
<td>2.540</td>
<td>2.520</td>
</tr>
<tr>
<td></td>
<td>(24.898)</td>
<td>(24.608)</td>
<td>(22.804)</td>
<td>(22.654)</td>
</tr>
<tr>
<td>LBP</td>
<td>1.230</td>
<td>1.264</td>
<td>0.776</td>
<td>0.826</td>
</tr>
<tr>
<td>PRICE</td>
<td>-0.365</td>
<td>-0.279</td>
<td>-1.266</td>
<td>-1.065</td>
</tr>
<tr>
<td></td>
<td>(-12.769)</td>
<td>(-8.213)</td>
<td>(-17.941)</td>
<td>(-12.104)</td>
</tr>
<tr>
<td>RPRICE-PRIICE</td>
<td>0.126</td>
<td></td>
<td>0.299</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(4.394)</td>
<td></td>
<td>(4.101)</td>
<td></td>
</tr>
<tr>
<td>PROMO</td>
<td>1.464</td>
<td>1.447</td>
<td>1.960</td>
<td>1.937</td>
</tr>
<tr>
<td>Purchase Incidence</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CONSUME</td>
<td>0.068</td>
<td>0.068</td>
<td>0.122</td>
<td>0.122</td>
</tr>
<tr>
<td></td>
<td>(34.628)</td>
<td>(34.963)</td>
<td>(23.281)</td>
<td>(23.334)</td>
</tr>
<tr>
<td>INVENTORY</td>
<td>-0.073</td>
<td>-0.074</td>
<td>-1.382</td>
<td>-1.383</td>
</tr>
<tr>
<td></td>
<td>(-2.676)</td>
<td>(-2.675)</td>
<td>(-21.295)</td>
<td>(-21.265)</td>
</tr>
<tr>
<td>CV</td>
<td>0.429</td>
<td>0.418</td>
<td>0.382</td>
<td>0.379</td>
</tr>
<tr>
<td></td>
<td>(12.456)</td>
<td>(12.245)</td>
<td>(11.680)</td>
<td>(13.511)</td>
</tr>
<tr>
<td>Log-likelihood</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BIC</td>
<td>-16924.600</td>
<td>-16914.391</td>
<td>-11949.074</td>
<td>-11940.608</td>
</tr>
<tr>
<td>In-sample</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hit-Rate (Choice)</td>
<td>54.82%</td>
<td>55.53%</td>
<td>57.23%</td>
<td>57.23%</td>
</tr>
<tr>
<td>Hit-Rate (Incidence)</td>
<td>91.05%</td>
<td>91.05%</td>
<td>94.42%</td>
<td>94.42%</td>
</tr>
</tbody>
</table>

a Brand-specific constants are not presented.
b All parameters are significant at $\alpha =0.05$
c Asymptotic $t$-statistics in parentheses.
d Log-likelihood and BIC are evaluated at the maximum likelihood estimates.
Table 2.9: HB-ST Parameter Estimates and Fit Statistics

<table>
<thead>
<tr>
<th>Variables</th>
<th>$\mu$ (Yogurt)</th>
<th>$\mu$ (Ketchup)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Brand Choice</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BLOY</td>
<td>2.476</td>
<td>3.315</td>
</tr>
<tr>
<td></td>
<td>(2.214, 2.738)</td>
<td>(2.886, 3.743)</td>
</tr>
<tr>
<td>LBP</td>
<td>0.324</td>
<td>0.16</td>
</tr>
<tr>
<td></td>
<td>(0.167, 0.481)</td>
<td>(0.025, 0.296)</td>
</tr>
<tr>
<td>PRICE</td>
<td>-0.482</td>
<td>-1.522</td>
</tr>
<tr>
<td></td>
<td>(-0.560, -0.404)</td>
<td>(-1.633, -1.410)</td>
</tr>
<tr>
<td>RPRICE-PRICE</td>
<td>0.052</td>
<td>0.135</td>
</tr>
<tr>
<td></td>
<td>(-0.024, 0.128)</td>
<td>(-0.038, 0.307)</td>
</tr>
<tr>
<td>PROMO</td>
<td>1.409</td>
<td>2.163</td>
</tr>
<tr>
<td></td>
<td>(1.223, 1.595)</td>
<td>(2.022, 2.303)</td>
</tr>
<tr>
<td><strong>Purchase Incidence</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CONSUME</td>
<td>0.098</td>
<td>0.166</td>
</tr>
<tr>
<td></td>
<td>(0.072, 0.124)</td>
<td>(0.133, 0.200)</td>
</tr>
<tr>
<td>INVENTORY</td>
<td>-0.524</td>
<td>-2.421</td>
</tr>
<tr>
<td></td>
<td>(-0.643, -0.407)</td>
<td>(-2.693, -2.148)</td>
</tr>
<tr>
<td>CV</td>
<td>0.447</td>
<td>0.360</td>
</tr>
<tr>
<td></td>
<td>(0.374, 0.520)</td>
<td>(0.292, 0.428)</td>
</tr>
<tr>
<td>Log-likelihood</td>
<td>-12663.221</td>
<td>-9832.465</td>
</tr>
<tr>
<td><strong>In-sample</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hit-Rate (Choice)</td>
<td>74.36%</td>
<td>74.35%</td>
</tr>
<tr>
<td>Hit-Rate (Incidence)</td>
<td>91.59%</td>
<td>94.50%</td>
</tr>
</tbody>
</table>

a Brand-specific constants are not presented.
b 95% probability interval in parentheses
c Log-likelihood is evaluated at the posterior estimates of household parameters ($\theta^h$)
Table 2.10: Regression of CV Parameters on Price Parameters

<table>
<thead>
<tr>
<th>Variables</th>
<th>Parameter Estimates</th>
<th>t-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Yogurt</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>0.335</td>
<td>8.562</td>
</tr>
<tr>
<td>Price Parameters</td>
<td>−0.233</td>
<td>−3.358</td>
</tr>
<tr>
<td><strong>Ketchup</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>0.184</td>
<td>3.492</td>
</tr>
<tr>
<td>Price Parameters</td>
<td>−0.117</td>
<td>−3.472</td>
</tr>
</tbody>
</table>
Table 2.11: Reference Price Elasticity for the Yogurt Category

<table>
<thead>
<tr>
<th>Brand</th>
<th>NL-ST model</th>
<th>HB-ST model</th>
<th>Percent Decrease</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yoplait</td>
<td>1.02</td>
<td>0.41</td>
<td>60%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-0.27, 1.10)</td>
<td></td>
</tr>
<tr>
<td>Weight Watcher</td>
<td>0.86</td>
<td>0.35</td>
<td>59%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-0.23, 0.93)</td>
<td></td>
</tr>
<tr>
<td>Dannon</td>
<td>0.94</td>
<td>0.37</td>
<td>61%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-0.24, 0.97)</td>
<td></td>
</tr>
<tr>
<td>Nordica</td>
<td>0.63</td>
<td>0.22</td>
<td>65%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-0.14, 0.58)</td>
<td></td>
</tr>
<tr>
<td>QC</td>
<td>0.56</td>
<td>0.21</td>
<td>62%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-0.14, 0.55)</td>
<td></td>
</tr>
<tr>
<td>W.B.B</td>
<td>0.56</td>
<td>0.19</td>
<td>66%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-0.12, 0.50)</td>
<td></td>
</tr>
<tr>
<td>Private Label</td>
<td>0.43</td>
<td>0.14</td>
<td>67%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-0.09, 0.37)</td>
<td></td>
</tr>
</tbody>
</table>

* a 95% probability interval in parentheses
Table 2.12: **Reference Price Elasticity for the Ketchup Category**

<table>
<thead>
<tr>
<th>Brand</th>
<th>NL-ST model</th>
<th>HB-ST model</th>
<th>Percent Decrease</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hunt's 32 Oz.</td>
<td>0.85</td>
<td>0.35</td>
<td>59%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-0.19, 0.88)</td>
<td></td>
</tr>
<tr>
<td>Del Monte 32 Oz.</td>
<td>0.95</td>
<td>0.40</td>
<td>68%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-0.21, 1.01)</td>
<td></td>
</tr>
<tr>
<td>Heinz 28 Oz.</td>
<td>0.98</td>
<td>0.38</td>
<td>61%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-0.21, 0.98)</td>
<td></td>
</tr>
<tr>
<td>Heinz 32 Oz.</td>
<td>0.68</td>
<td>0.25</td>
<td>63%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-0.13, 0.64)</td>
<td></td>
</tr>
<tr>
<td>Heinz 40 Oz.</td>
<td>1.28</td>
<td>0.55</td>
<td>57%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-0.29, 1.38)</td>
<td></td>
</tr>
<tr>
<td>Heinz 64 Oz.</td>
<td>1.28</td>
<td>0.54</td>
<td>58%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-0.29, 1.37)</td>
<td></td>
</tr>
<tr>
<td>Private Label</td>
<td>0.78</td>
<td>0.32</td>
<td>59%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-0.17, 0.82)</td>
<td></td>
</tr>
</tbody>
</table>

A 95% probability interval in parentheses
Table 2.13: Price Elasticity for the Yogurt Category

<table>
<thead>
<tr>
<th>Brand</th>
<th>NL-ST model</th>
<th>HB-ST model</th>
<th>Percent Increase</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yoplait</td>
<td>-3.33</td>
<td>-4.73</td>
<td>142%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-5.48, -3.98)</td>
<td></td>
</tr>
<tr>
<td>Weight Watcher</td>
<td>-2.76</td>
<td>-3.99</td>
<td>145%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-4.60, -3.38)</td>
<td></td>
</tr>
<tr>
<td>Dannon</td>
<td>-3.04</td>
<td>-4.17</td>
<td>137%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-4.82, -3.52)</td>
<td></td>
</tr>
<tr>
<td>Nordica</td>
<td>-2.05</td>
<td>-2.55</td>
<td>124%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-2.91, -2.19)</td>
<td></td>
</tr>
<tr>
<td>QC</td>
<td>-1.80</td>
<td>-2.36</td>
<td>131%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-2.72, -2.00)</td>
<td></td>
</tr>
<tr>
<td>W.B.B</td>
<td>-1.82</td>
<td>-2.18</td>
<td>120%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-2.50, -1.87)</td>
<td></td>
</tr>
<tr>
<td>Private Label</td>
<td>-1.39</td>
<td>-1.61</td>
<td>116%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-1.86, -1.36)</td>
<td></td>
</tr>
</tbody>
</table>

a 95% probability interval in parentheses
Table 2.14: Price Elasticity for the Ketchup Category

<table>
<thead>
<tr>
<th>Brand</th>
<th>NL-ST model</th>
<th>HB-ST model</th>
<th>Percent Increase</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hunt's 32 Oz.</td>
<td>-3.95</td>
<td>-4.76</td>
<td>121%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-5.34, -4.18)</td>
<td></td>
</tr>
<tr>
<td>Del Monte 32 Oz.</td>
<td>-4.38</td>
<td>-5.36</td>
<td>122%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-5.95, -4.77)</td>
<td></td>
</tr>
<tr>
<td>Heinz 28 Oz.</td>
<td>-4.54</td>
<td>-5.29</td>
<td>117%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-5.89, -4.68)</td>
<td></td>
</tr>
<tr>
<td>Heinz 32 Oz.</td>
<td>-3.16</td>
<td>-3.44</td>
<td>109%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-3.83, -3.05)</td>
<td></td>
</tr>
<tr>
<td>Heinz 40 Oz.</td>
<td>-5.90</td>
<td>-7.34</td>
<td>124%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-8.20, -6.48)</td>
<td></td>
</tr>
<tr>
<td>Heinz 64 Oz.</td>
<td>-5.83</td>
<td>-7.18</td>
<td>123%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-8.02, -6.33)</td>
<td></td>
</tr>
<tr>
<td>Private Label</td>
<td>-3.59</td>
<td>-4.35</td>
<td>121%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-4.83, -3.87)</td>
<td></td>
</tr>
</tbody>
</table>

* a 95% probability interval in parentheses
Chapter 3

Consideration Set Heterogeneity, Advertising, and Price Sensitivity

3.1 Overview

Recently, several models have been introduced in the marketing literature to understand better the role of choice or consideration sets in brand choice behavior. These models view brand choice as a two-step process: a consideration set is formed in the first stage and, in the second stage, consumers evaluate the brands in the consideration set. The use of consideration sets by consumers can be theoretically justified in two ways. First, consumers may choose not to evaluate some brands for which they have a low preference or have high evaluation costs (e.g., Hauser and Wernerfelt 1990, Roberts and Lattin 1991). Second, even if evaluation costs are low, consumers may not have the ability to process all the information available to them. Consequently, they may simplify the decision process by only evaluating a subset of alternatives (Meyer and Kahn 1991). The resulting consideration set is a product of a simple non-compensatory, process rather than a more elaborate, compensatory, process (e.g., Lehmann and Pan 1994, Shocker et. al. 1991).

In this chapter, we develop a new consideration set model, which can be estimated with scanner panel data, to account better for heterogeneity in consideration sets. The proposed approach directly models the probability of a brand being included in the consideration set. The resulting inclusion probabilities for brands reflect a "fuzzy" consideration set, i.e., the composition of the consideration set is fuzzy in the sense that a
brand belongs to the consideration set only probabilistically. This approach differs from that of most previous research (e.g., Siddarth, Bucklin and Morrison 1995, Andrews and Srinivasan 1995, Chiang, Chib and Narasimhan 1998), in which consideration sets are well-defined and the researcher models the uncertainty about the set from which a brand is chosen. Our approach has a modeling advantage and better represents some findings from the consumer behavior literature (Shocker et al. 1991) as discussed below.

The approach also differs from the fuzzy set model of Bronnenberg and Vanhonacker [B&V] (1996) in which variables have a compensatory effect on both brand evaluation and consideration. In contrast, our model permits a non-compensatory process in consideration set formation while allowing brand evaluation to be compensatory. We apply the proposed fuzzy set model and find that it outperforms several alternative consideration set models (in two product categories) proposed in the literature.

We apply our approach to examine the role of the consideration set in moderating the impact of advertising on price sensitivity. This is a generalization of Mitra and Lynch (1995) — using real purchase data as opposed to experimental data — that overcomes several limitations of their experimental approach.

We seek to make the following contributions.

- Develop a new fuzzy consideration set model and compare its performance to existing consideration set models.

- Apply the proposed model to investigate the role of the consideration set in moderating the effect of advertising on price sensitivity.

In the remainder of this section, we review existing approaches to incorporating consideration sets in choice models and highlight the methodological differences of the proposed approach. We provide a brief review of previous work on the relationship between advertising and price sensitivity, and discuss the manner in which our approach overcomes
some of the limitations of prior research in this.

The remaining sections of this chapter are organized as follows. Section 3.2 develops the new fuzzy consideration set model and provides mathematical specifications of the model components. Section 3.3 extends the model to include the impact of advertising on price sensitivity through its influence on the consideration set. Section 3.4 presents an empirical application of our model to yogurt and ketchup scanner panel data sets. Section 3.5 discusses the empirical results and Section 3.6 presents the managerial implications of the study. We close with our conclusions and directions for future research.

3.1.1 Approaches to Incorporating Consideration Sets in Choice Models

Non-Fuzzy Set Approaches

One approach to modeling consideration sets is, what we term as, the Universal Set Approach (UNSA) (e.g., Andrews and Srinivasan 1995, Chiang, Chib and Narasimhan 1998), in which all possible consideration sets are enumerated and the associated probability of a consumer using each set is explicitly modeled. As a result, as the number of alternatives (n) increases, the number of consideration sets ($2^n - 1$) increases exponentially as does the difficulty in model estimation.

Another approach, which we term as the Restricted Set Approach (RSA), circumvents this problem by imposing *a priori* restrictions on the number of subsets used by the consumer (e.g., Siddarth, Bucklin and Morrison 1995, Gaudry and Dagenais 1979, Swait and Ben-Akiva 1987). However, the RSA cannot avoid mis-specifying consideration sets. For example, a chosen brand may not be in the consideration set or the actual consideration set may be smaller than the specified consideration set. A feature shared by both approaches is that consideration sets are well-defined, the consideration sets are "crisp" as termed in Elrod et al. (1992) and only the probability associated with the use
Fuzzy Set Approaches

A "Fuzzy" set approach directly models the probability of an individual brand being included in the consideration set (Fotheringham 1988, Bronnenberg and Vanhonacker [B&V] 1996). The composition of the consideration set is "fuzzy", as compared to "crisp", in the sense that a brand belongs to the consideration set only probabilistically. This contrasts with the use of the non-fuzzy consideration sets in which the sets are well-defined but there is uncertainty about the set from which a brand is chosen.

One distinct advantage of the fuzzy set approach is that it can easily accommodate many possible consideration sets. Consider a \( n \) brand market. Let \( \delta_i \) indicate whether brand \( i \) is, or is not, included in the consideration set. The UNSA permutes the \( \delta_i \) while restricting its value to either 1 or 0 and thus the total number of possible consideration sets, excluding the empty set, is \( (2^n - 1) \). On the other hand, the fuzzy set approach allows the value of \( \delta_i \) to lie between 0 and 1, defining infinitely many consideration sets in a single fuzzy set. This approach requires the use of \( n \) brand-specific inclusion probabilities as opposed to the \( 2^n - 1 \) set-specific probabilities in the UNSA. In addition, the fuzzy set approach is consistent with the behavioral reality that consumers themselves may be unable to disclose the contents of their consideration sets with certainty (Shocker et al. 1991).

B&V (1996) model consideration via a fuzzy set approach. They use a brand-specific binary logit expression (based on a linear additive utility function) to model its inclusion probability, which results in a compensatory model of consideration set formation. In contrast, we hypothesize a non-compensatory consideration process and, therefore, use a Multiplicative Competitive Interaction (MCI) model to specify inclusion probabilities. The two approaches represent two different conceptualizations of consideration set
information process which we compare later on both theoretical and empirical criteria.

3.1.2 Advertising, Price Sensitivity and Consideration Sets

There are two, conflicting views on the effect of advertising on price sensitivity. The "Market Power" hypothesis contends that advertising decreases price sensitivity by creating brand loyalty, reducing competition and thus building market power (Comanor and Wilson 1979). The alternative, "Information", hypothesis is that advertising provides information which makes consumers aware of more alternatives and thus become more price-sensitive (Nelson 1974).

Numerous empirical studies that have examined this relationship have produced contradictory findings. Lambin (1976) regressed the absolute value of price elasticities on several measures of advertising intensity and found a significant inverse relationship between ad expenditure and price elasticity, consistent with the market power hypothesis. In contrast, Wittink (1977) found that territories with higher advertising levels were more price-sensitive, consistent with the information hypothesis. Prasad and Ring (1976) investigated the interaction effects of advertising and price on market shares at two controlled levels of aggregate TV advertising exposure and found a positive relationship between advertising and price sensitivity. Krishnamurthi and Raj (1985) compared sales in a test period with high advertising to a pretest period with normal advertising. Both control and experimental panels were classified into high and low price-sensitive groups. The high price-sensitive group in the experimental panel was found to become less price-sensitive, supporting the market power hypothesis. Kanetkar, Weinberg and Weiss (1992), using scanner panel data, found that increased exposure to advertising was associated with an increase in price sensitivity, supporting the information hypothesis. In contrast, Papatla (1995), also using scanner panel data, showed that advertising reduced price sensitivity.

Several studies have attempted to resolve the conflicts in previous empirical findings.
Boulding, Lee and Staelin (1994) speculate that the content of advertising might provide a potential explanation. For example, price-oriented advertising by retailers may increase price elasticity, and image-oriented brand advertising by manufacturers may decrease price elasticity. Gatignon (1984) proposes that competitive reaction to own brand's advertising is an important moderator of its effect on price sensitivity. Therefore, even if the direct effect of brand advertising is to build market power, it may trigger competitive counter-advertising (competitive increase in advertising expenditures) that may make the market become more aware of alternatives and, hence, more price-sensitive.

Mitra and Lynch (1995) highlight the role that consideration sets may play in moderating the impact of advertising on price elasticity. In their experimental study, they hypothesize the following process at work. First, advertising increases price elasticity by expanding consideration set size ("Information"). Second, advertising decreases price elasticity by increasing relative strength of preference ("Market Power"). Third, increased relative strength of preference may also reduce consideration set size, decreasing price elasticity ("Market Power"). The net effect of advertising on price elasticity, therefore, depends on the combined impact of these processes. Two pretests establish the hypothesized relationship between advertising and consideration set size and that between advertising and relative strengths of preference. The main experiment varies the relative strength of each relationship via different choice environments (memory-based and stimulus-based) and advertising conditions (reminder vs. differentiating advertising), and shows that the direction of the effect of advertising on price elasticity depends on the strength of these relationship. Thus, their framework provides a unifying explanation of advertising effects on price elasticity by explicitly accounting for the moderating role of the consideration set. What follows from their conceptual model is that the net observed effect of advertising may support either the "Market Power" or the "Information" theory, but not both, if the separate effects on consideration set formation are ignored.
In general, previous research in this area is limited in several ways. First, as is often the case in econometric studies, most of the studies have been performed at various level of aggregation without apparent control for possible aggregation bias (Gatignon 1984, Kanetkar et al. 1992, and Papatla 1995 are exceptions.). Second, advertising expenditures were most often used as proxy measure for advertising exposure (Kanetkar et al. 1992 and Papatla 1995 are exceptions.). More importantly, these studies did not account for the consideration set that is hypothesized to moderates the effect of advertising (Mitra and Lynch 1995).

While the results reported in Mitra and Lynch (1995) supported their hypotheses, their experimental study was limited in two important ways. First, to establish the direction of causality, they compared the case in which all brands advertised to the case in which none of the brands engaged in advertising. Specifically, all subjects in the experimental group are assumed to be exposed to the same level of advertising. Thus, household heterogeneity in advertising exposure levels is not controlled for in the analysis. Consequently, as Popkowski-Leszczyc and Rao (1989) suggest, household differences in advertising susceptibility may lead to biased estimates. (This may also be true of other previous studies that have not accounted for heterogeneity in consumer responses to advertising). Second, to control for prior brand preference (to minimize the effects of prior preference for the brand and to maximize the power to detect the effect of advertising), unfamiliar brands were used in the experiment. Therefore, increased advertising for all brands could only increase the consideration set size. Our current study seeks to generalize their findings using a more natural competitive, field, setting that overcomes the above limitations of their experiment.
3.2 Proposed Fuzzy Set Model

The consideration set evolves from the awareness set and "can be viewed as consisting of alternatives salient or accessible on a particular occasion" (Shocker et al. 1991, p.183). Some brands may be more accessible from memory and therefore, included in the consumer's consideration set, e.g., brands that the consumer has purchased in the past or brands for which the consumer has recently been exposed to brand advertising. This initial consideration set may change as a result of in-store stimuli such as displays or point-of-purchase (POP) materials. In summary, the consideration set is a dynamic construct that varies across consumers, within consumers across purchase occasions, and even within a given purchase occasion (Nedungadi 1990, Shocker et. al. 1991). This aspect guides the choice of variables that influence consideration.

Another important aspect of consideration set formation reflected in our modeling approach is its non-compensatory nature (Lehmann and Pan 1994, Shocker et. al. 1991). Consumers are assumed to screen out brands in a non-compensatory manner to form the consideration set, and then to evaluate only a subset of brands in a compensatory manner. We implement a fuzzy version of the non-compensatory rule in which every brand has a non-zero probability of being included in the consideration set.

3.2.1 Modeling Approach

We assume that overall brand utility is affected by variables that influence brand evaluation and the consideration set. Let $X_{ht}^i$ and $Y_{ht}^i$ be vectors of variables that affect brand evaluation and consideration of household $h$ for brand $i$ on purchase occasion $t$. The intersection of the two vectors may be non-empty. For example, brand advertising may influence the consideration set, by making the brand more accessible in memory, and also brand evaluation, by distinguishing it from other brands in the consideration set.
Chapter 3. Consideration Set Heterogeneity, Advertising, and Price Sensitivity

We begin by specifying the deterministic component of the overall utility, $OU^h_{it}$, as follows:

$$OU^h_{it} = \sum_i \beta_i f_x(x^h_{it}) + \sum_m \gamma_m f_y(y^h_{imt}),$$  \hspace{1cm} (3.1)

where $\beta_i$ and $\gamma_m$ are parameters for brand evaluation and consideration variables, respectively, and, $f_x$ and $f_y$ denote functional relationships between each set of variables and the overall utility. This is a general formulation and, in order to develop the model further, we must choose appropriate functional forms for $f_x$ and $f_y$.

Impact of $X^h_{it}$

For the moment, let us assume that household $h$ considers all available brands and starts evaluating brands at the point of purchase. Further, ignoring $Y^h_{it}$ and using an identity transformation function for $f_x$, we can specify the utility function as:

$$U^h_{it} = \sum_i \beta_i x^h_{it},$$  \hspace{1cm} (3.2)

where $U^h_{it}$ denotes the deterministic component of the conventional utility function. This represents the typical multinomial logit (MNL) brand choice model in which it is assumed that all brands are considered and evaluated in a compensatory manner. Since exact utilities are unknown, we can also add a random error term, $\epsilon_{it}$. Then the probability of brand $i$ being chosen is given by:

$$P_i^h = p[U^h_{it} + \epsilon_{it} > U^h_{kt} + \epsilon_{kt}(k \in K, k \neq i)]$$  

$$= p[\epsilon_{kt} < \epsilon_{it} + U^h_{it} - U^h_{kt}(k \in K, k \neq i)].$$  \hspace{1cm} (3.3)

\footnote{We may either omit $Y^h_{it}$ from the model or include it in $X^h_{it}$, assuming that all variables share the same functional relationship, $f_x$, to the utility.}
Without making any assumption about the distribution of the error terms, McFadden (1974) shows that Equation (3.3) may be written as:

\[ P_t^h(i) = \int_{x=-\infty}^{\infty} g(\varepsilon_{it} = x) \prod_{k \neq i}^{K} \int_{y=-\infty}^{x+U_{it}^h - U_{kt}^h} g(\varepsilon_{kt} = y) dy dx, \]  

(3.4)

where \( g(\cdot) \) represents the probability density function of the error.

However, in general, consumers do not evaluate all \( K \) alternatives but instead only evaluate a subset, \( C \), of them where \( C \in K \) and, from either an analyst's or a consumer's point of view, there is uncertainty about the composition of \( C \). Therefore, the set \( C \) can be considered to be "fuzzy" (Fotheringham 1988). This uncertainty can be incorporated by rewriting Equation (3.3) as:

\[ P_t^h(i) = p [pr_t^h(i \in C)(U_{it}^h + \varepsilon_{it}) > pr_t^h(k \in C)(U_{kt}^h + \varepsilon_{kt})(k \in K, k \neq i)], \]  

(3.5)

where the utility for brand \( i \) is rescaled by \( pr_t^h(i \in C) \), the probability of brand \( i \) being included in the consideration set \( C \). Assuming that \( \varepsilon_{it} \) is independently and identically distributed with a Type I extreme value distribution but making no assumption about the values of the \( pr_t^h(i \in C) \) terms (Fotheringham 1988), Equation (3.5) can be expressed as follows:

\[ P_t^h(i) = pr_t^h(i \in C) \int_{x=-\infty}^{\infty} g(\varepsilon_{it} = x) \prod_{k \neq i}^{K} pr_t^h(k \in C) \int_{y=-\infty}^{x+U_{it}^h - U_{kt}^h} g(\varepsilon_{kt} = y) dy dx \]

\[ = \int_{x=-\infty}^{\infty} \exp(-x) \prod_{k \neq i}^{K} pr_t^h(k \in C) \exp[- \exp(-x - U_{it}^h + U_{kt}^h)] dx \]

\[ = \frac{pr_t^h(i \in C) \exp(U_{it}^h)}{\sum_k pr_t^h(k \in C) \exp(U_{kt}^h)}. \]  

(3.6)

For notational simplicity and ease of understanding, we term \( pr_t^h(i \in C) \) as the inclusion probability, \( IP_t^h \). In the subsequent discussion, we provide a specific functional form for \( IP_t^h \).
Impact of $Y^h_{it}$

We adopt a fuzzy version of the non-compensatory rule in which every brand has a non-zero probability of being included in the consideration set. We specify the inclusion probability, $IP^h_{it}$, as a function of $Y^h_{it}$, using a Multiplicative Competitive Interaction (MCI) functional form. Thus, we have:

$$IP^h_{it} = \frac{\prod_m (y^h_{int})^{\gamma_m}}{\sum_k \prod_m (y^h_{kmt})^{\gamma_m}}$$

(3.7)

Using an MCI form for $IP^h_{it}$ is equivalent to choosing $f_y$ different from $f_x$ in the overall utility expression in Equation (3.1). The following derivation makes this point clear:

$$\frac{IP^h_{it} \exp(U^h_{it})}{\sum_k IP^h_{kt} \exp(U^h_{kt})} = \frac{\prod_m (y^h_{int})^{\gamma_m} \exp(U^h_{it})}{\sum_k \prod_m (y^h_{kmt})^{\gamma_m} \exp(U^h_{kt})}$$

$$= \frac{\prod_m (y^h_{int})^{\gamma_m} \exp(U^h_{it})}{\sum_k \prod_m (y^h_{kmt})^{\gamma_m} \exp(U^h_{kt})}$$

$$= \frac{\exp[\ln(\prod_m (y^h_{int})^{\gamma_m})] \exp(U^h_{it})}{\sum_k \exp[\ln(\prod_m (y^h_{kmt})^{\gamma_m})] \exp(U^h_{kt})}$$

$$= \frac{\exp[\sum_m \ln (y^h_{int})^{\gamma_m}] \exp(U^h_{it})}{\sum_k \exp[\sum_m \ln (y^h_{kmt})^{\gamma_m}] \exp(U^h_{kt})}$$

$$= \frac{\exp[U^h_{it} + \sum_m \gamma_m \ln(y^h_{int})]}{\sum_k \exp[U^h_{kt} + \sum_m \gamma_m \ln(y^h_{kmt})]}$$

$$= \frac{\exp[\sum_i \beta_i x^h_{it} + \sum_m \gamma_m \ln(y^h_{int})]}{\sum_k \exp[\sum_i \beta_i x^h_{kt} + \sum_m \gamma_m \ln(y^h_{kmt})]}$$

(3.8)

In other words, specifying an MCI form for $IP^h_{it}$ is, in effect, equivalent to using log-transformed $Y^h_{it}$ in the overall utility function as opposed to the identity transformation used for $X^h_{it}$. By allowing the functional forms to differ (Roberts and Lattin 1997), the analyst can capture the non-compensatory nature of consideration set formation while allowing brand evaluations to be compensatory (e.g., Bettman 1979; Gensch 1987). In addition, this feature of model formulation makes it possible to identify separately\(^2\) the

\(^2\)In practice, however, this formulation may suffer from multi-collinearity problem as reported in Andrews and Srinivasan (1995).
Chapter 3. Consideration Set Heterogeneity, Advertising, and Price Sensitivity

effects of the same variable used in consideration and brand evaluation, i.e., when the intersection of $X^h_{it}$ and $Y^h_{it}$ is non-empty (This feature distinguishes our model from the B&V model, which we will discuss later).

In effect, the log-transformation helps screen out brands that are unsatisfactory on any one attribute ($y^h_{imt}$) of $Y^h_{it}$. To illustrate this point, we present functional shapes that reflect a compensatory process, a fuzzy version and a crisp version of a non-compensatory process, respectively, in Figure 3.1. The fuzzy non-compensatory line (log function) shows that the utility begins to fall off rapidly for alternatives with $y^h_{imt}$ close to zero. This happens when brand preference (conventionally captured via a brand loyalty measure that lies between 0 and 1), for example, in $Y^h_{it}$ comes close to zero, resulting in predicted choice probability close to zero. The utility for a brand with negligible preference cannot be compensated for by other attributes. In contrast, the compensatory line reflects that the utility can be compensated for by other attributes even if $y^h_{imt}$ comes to zero. As compared to the fuzzy non-compensatory line, the crisp non-compensatory line describes a kinked utility function that allows for a immediate fall-off for $y^h_{imt}$ less than a certain threshold (Elrod et al. 1992).

---

3The non-compensatory process adopted in the current study is based on the conjunctive decision rule described in Elrod et al. 1992. Unlike the lexicographic-type Elimination-By-Aspects model (Tversky 1972) that screens alternatives on an attribute-by-attribute basis in some hierarchical sequence, our conjunctive decision rule does not assume a sequential consideration of attributes but a parallel consideration of attributes along with attribute weights. Therefore, among a predetermined set of consideration attributes, if any one attribute level of a brand comes close to zero, that brand is likely to be screened out.

4In effect, uncertainty about the composition of the consideration set arises from fuzziness of the threshold level, and is represented by smoothness of the utility function implied by a fuzzy version of non-compensatory process.
Comparison to B&V (1996)

Bronnenberg and Vanhonacker [B&V] (1996) define the inclusion probability for each brand using an independent binary logit expression as follows:

\[
IP^h_{it} = \frac{\exp(\gamma Y^h_{it})}{\exp(\gamma Y^h_{it}) + \exp(\Delta_t)} \tag{3.9}
\]

where \(\Delta_t\) is a threshold (to be estimated) against which inclusion utility, \(\gamma Y^h_{it}\), is compared.

To compare the two approaches, we use the above Equation (3.9) to substitute for \(IP^h_{it}\) in Equation (3.8) as follows:

\[
\frac{IP^h_{it} \exp(U^h_{it})}{\sum_k IP^h_{kt} \exp(U^h_{kt})} = \frac{\exp(\gamma Y^h_{it})[\exp(\gamma Y^h_{it}) + \exp(\Delta_t)]^{-1} \exp(U^h_{it})}{\sum_k \exp(\gamma Y^h_{kt})[\exp(\gamma Y^h_{kt}) + \exp(\Delta_t)]^{-1} \exp(U^h_{kt})}
\]

\[
= \frac{\exp\left[U^h_{it} + \gamma Y^h_{it} - \ln[\exp(\gamma Y^h_{it}) + \exp(\Delta_t)]\right]}{\sum_k \exp\left[U^h_{kt} + \gamma Y^h_{kt} - \ln[\exp(\gamma Y^h_{kt}) + \exp(\Delta_t)]\right]}
\]

\[
= \frac{\exp\left[\sum_i \beta_i x^h_{iit} + \sum_m \gamma_m y^h_{itmt} - \ln[\exp(\sum_m \gamma_m y^h_{itmt}) + \exp(\Delta_t)]\right]}{\sum_k \exp\left[\sum_i \beta_i x^h_{kit} + \sum_m \gamma_m y^h_{kmt} - \ln[\exp(\sum_m \gamma_m y^h_{kmt}) + \exp(\Delta_t)]\right]} \tag{3.10}
\]

Therefore, in contrast to our model, in B&V (1996), the \(Y^h_{it}\) appear in the overall utility function after an identity transformation. This implies that in the B&V model, both brand evaluation and consideration variables have a compensatory influence on overall utility and unlike in our model, explicit screening based on \(\{y^h_{itmt}\}\) does not take place. Note that, in the B&V model, due to identification problems, the same variable cannot influence both brand evaluation and consideration. In contrast, our model permits us to separately identify the impact of a variable on both stages.
3.3 Advertising, Price Sensitivity and Consideration Sets

In generalizing Mitra and Lynch (1995) in the current study, we focus on the direct effect\(^5\) of advertising on consideration set size. Unlike their experimental study, which hypothesizes that advertising increases consideration set size (that should be true when all unfamiliar brands advertise) and therefore increases price sensitivity, the current study hypothesizes the following process at work.

Advertising can either increase or decrease price sensitivity through its influence on the consideration set. First, brand advertising can increase inclusion probability for a brand. Second, consideration set size can either increase or decrease as a result of advertising. For example, if in response to advertising, the distribution of inclusion probabilities across brands becomes more even (skewed), then consideration set size increases (decreases). Third, as the consideration size increases (decreases), the consumer perceives brands as more (less) close substitutes, becoming more (less) price-sensitive. This indirect relationship between advertising and price sensitivity is presented in Figure 3.2.

We examine the first link by specifying the inclusion probability as a function of advertising and the second link by measuring consideration set size via the entropy of inclusion probabilities. Gensch and Soofi (1994) illustrate a mapping of an information index\(^6\) to consideration set size by using linear interpolation. The logic is that in the absence of attribute information, predicted choice probabilities are uniform across brands (all brands are considered), resulting in maximum entropy (maximum uncertainty), \(\ln(K)\), where \(K\) = the number of brands. If attribute information is available, then predicted

\(^5\)Mitra and Lynch (1995) also investigates the indirect effect of advertising on consideration set size through its influence on relative strength of brand preference. Since we do not model brand preference as a function of advertising, but infer it from revealed choice history, we focus on the direct effect of advertising on consideration set size.

\(^6\)The information index is defined as \(1 - \frac{\sum_{k=1}^{K} \pi_k \ln \pi_k}{\ln K}\) where \(\pi_k\) is the estimated choice probability, \(\ln K\) is the maximum entropy and \(K\) is the number of brands. The information index is negatively related to a set size.
Chapter 3. Consideration Set Heterogeneity, Advertising, and Price Sensitivity

choice probabilities are pulled away from the uniformity, resulting in reduced entropy which reflects the reduction of the number of choices in a decision problem from \( K \) to less than \( K \). Following the logic, we use the ratio of the entropy of inclusion probabilities to the maximum entropy as a measure of a consideration set size since it provides more complete, continuous, information as opposed to discrete approximation of a set size. In addition, a continuous measure of set size is consistent with our view that the composition of the consideration set is fuzzy rather than crisp. The relative entropy for household \( h \) on occasion \( t \) is calculated as follows:

\[
\text{ENT}_t^h = \frac{\sum_k -\text{IP}^h_{kt} \ln(\text{IP}^h_{kt})}{\ln(K)},
\]

where \( \text{ENT}_t^h \) is the relative entropy for household \( h \) on occasion \( t \) and \( K \) is the number of brands. The relative entropy takes the maximum value of 1 when all inclusion probabilities are equal (consideration set size of \( K \)) and the minimum value of 0 when any one brand has an inclusion probability of one (consideration set size of one) and all other brands have inclusion probabilities of zero. Therefore, the relative entropy is positively related to a consideration set size.

To investigate the last link, we use a varying parameter approach (Gatignon and Hanssens 1987). This approach allows us to examine the relationship between the consideration set size and price sensitivity by reparameterizing the price parameter as a function of the relative entropy. In doing so, it is critical to account for differences in inherent price sensitivity to capture the correct relationship since consideration set size does not explain inherent price sensitivity. For example, consumers with the same consideration set size may show quite different price sensitivities depending on which brands are usually in the consideration set. If low price-tier brands are in the set, price sensitivity

\[\text{Alternatively, we could use the entropy of inclusion probabilities to derive a consideration set size by solving for } S \text{ in equation } \ln(S) = \sum_k -\text{IP}^h_{kt} \ln(\text{IP}^h_{kt}) \text{ where } S \text{ lies between } 1 \text{ and } K. \text{ We found that this alternative measure } (S) \text{ provided a equivalent fit in the empirical application.} \]
will be high, and vice versa. Therefore, we assume that price sensitivity has two components, inherent (time-invariant) price sensitivity and inter-temporal price sensitivity. In a regression context, differences in inherent price sensitivity are captured by separate intercepts across segments. Beyond inherent price sensitivity, inter-temporal variations in price sensitivity are captured by the relative entropy (consideration set size).

3.4 Empirical Application

3.4.1 Estimation

For notational simplicity, so far, we have developed the fuzzy set model in a single-segment framework in which consumers are assume to share common response parameters. In estimating the model, we employ a latent class approach to account for response heterogeneity in both stages (Bucklin and Gupta 1992, Kamakura and Russell 1989). Therefore, all parameters are allowed to vary across segments. This is particularly important in the current study since, in the absence of heterogeneity correction, response parameters for inclusion probabilities may be inflated by erroneously capturing unaccounted for preference and response heterogeneity in brand choice. In addition, a hypothesized relationship between advertising and price sensitivity, in particular, should be investigated after controlling for price response heterogeneity (Krishnamurthi and Raj 1985, Papatla 1995) and heterogeneity in advertising susceptibility (Popkowski-Leszczyc and Rao 1989)

\footnote{A homogeneous logit choice model is reported to overestimate the impact of household-specific loyalty measures since they capture unaccounted for heterogeneity (e.g. Siddarth et al. 1995).}
3.4.2 Data

The data for this study are drawn from A.C. Nielsen scanner panel records, in the yogurt and ketchup categories, for households in Sioux Falls, South Dakota, for the period 1986-1988. The last 51 weeks of the data were used for model calibration and the preceding 61 weeks were used for initializing model variables. Households were qualified for inclusion in the sample if they made at least one grocery purchase every four weeks over the entire study period and made at least one purchase both in initialization and calibration periods. 599 and 592 households qualified on this basis for the yogurt and ketchup categories, respectively.

A technical problem that arises with single-source data is missing information in the household commercial exposure data due to breaks in the connection between household TVs and telemeters. Since some of the ads that households are exposed to may go undetected, the observed ad exposure rates in the database may underestimate the true household exposure rates. Following the screening procedure of Pedrick and Zufryden (1991), we included only those purchase records of the qualified households in the study in which their telemeter was working for over 21 out of the 28 days prior to each occasion. Thus, in the yogurt category, out of the total of 5,813 purchase records, 4,446 purchase records were included in the study. Out of the qualified households, 400 were randomly selected for model calibration (3053 purchases) and 199 for model validation (1393 purchases). In the ketchup category, we chose not to include ad exposure data because only two brands out of seven advertised on TV during the study period. Out of the qualified households, 400 were randomly selected for model calibration (2476 purchases) and 192 for model validation (1140 purchases).

The top seven selling brands are included in the study. Together, these accounted for 84 (85) percent of category sales in dollars, and 89 (87) percent in units for the yogurt
and ketchup categories, respectively.

### 3.4.3 Model Variables

We now specify the model variables that influence brand evaluation and consideration.

**Brand Evaluation**

We assume that consumers choose a brand by trading off between price and quality among actively considered brands. We interpret the brand-specific intercepts in the utility function as quality scores that can be decomposed into brand attributes and weights for those attributes (Elrod and Keane 1995). Therefore, the deterministic component of utility for each brand $i$ for household $h$ in segment $s$ is specified as a function of a brand-specific constant and price (e.g., Elrod and Keane 1995, B&V 1996 and Erdem 1996):

$$ U_{st}^h = \beta_{0i} + \beta_{s1}\text{PRICE}_{it}, \quad (3.12) $$

where $\text{PRICE}_{it} =$ actual shelf price of brand $i$ at time $t$.

**Consideration Set Formation**

In the absence of self-reported consideration set information, we may infer the brands which consumers consider based on their brand preference as measured by their prior experience with the brand (Nedungadi 1990, Andrews and Srinivasan 1995, Siddarth et. al. 1995, Chiang et. al. 1998). In a similar spirit, B&V (1996) propose that inclusion probabilities are related, either positively (reinforcing) or negatively (variety-seeking), to recency of purchase.\(^9\) This variable assumes the value 1 for the brand last purchased and

\(^9\) When choice behavior is of the reinforcing type, recently purchased brands are more likely to be salient to consumers. In contrast, when choice behavior is of the variety-seeking type, recently purchased brands are less likely to be salient to consumers (B&V 1996).
exponentially slopes downward for other brands.

The consideration set is dynamic and can change until a time immediately prior to brand choice (Shocker et al. 1991). Therefore, our model includes variables that capture brand preference, recency, TV ads, feature, and display activity. Marketing communication variables, such as TV ads and feature ads may provide cues to aid retrieval of brands from memory for inclusion into the consideration set. Moreover, after a consumer is in the store, he/she may consider additional brands based on their display (e.g., Nedungadi 1990, Andrews and Srinivasan 1995, B&V 1996). In addition, we propose to test whether there is an interaction effect of brand preference and TV advertising on consideration. Tellis (1988b) has shown that the impact of advertising on the probability of brand choice is higher for brands with higher preference. An underlying assumption of his study is that advertising affects brand choice by directly increasing brand utility and, hence, the probability of brand choice. In contrast, the well-known Hierarchy of Effects model (Lavidge and Steiner 1961) predicts that advertising will have a greater impact at a lower level of the hierarchy, i.e., awareness, knowledge, liking, and preference.

In the current study, unlike Tellis (1988b), we investigate the interaction effects of advertising and brand preference on consideration set formation. It may be argued that advertising for the most preferred brand might not increase the inclusion probability since the brand is most likely to be already included in the consideration set, i.e., a ceiling effect. On the other hand, advertising may have a substantial impact on consumers who currently have a low inclusion probability for a brand. Ultimately, whether the impact of advertising on inclusion probability (consideration set formation) is higher for low preference (low market share) brands or for high preference (high market share) brands is an empirical question which can be tested with data.

\[10\] Even though the consideration stage is not clearly defined in the Hierarchy of Effects model, it can be reasonably assumed that consideration stage corresponds to a lower level of the hierarchy that precedes action stage (brand choice).
In summary, we specify the inclusion probability as a function of preference, recency of choice, TV ads, interaction between brand preference and TV ads, feature ads, and display:

$$I^h_{sit} = \frac{(AD^h_{it})^{(\gamma_{b0} + \gamma_{b1} \ln(\pi^h_t))} (\pi^h_t)^{\gamma_{b2}} (REC^h_{it})^{\gamma_{b3}} (FEAT_{it})^{\gamma_{b4}} (DISP_{it})^{\gamma_{b5}}}{\sum_k (AD^k_{it})^{(\gamma_{b0} + \gamma_{b1} \ln(\pi^k_t))} (\pi^k_t)^{\gamma_{b2}} (REC^k_{it})^{\gamma_{b3}} (FEAT_{it})^{\gamma_{b4}} (DISP_{it})^{\gamma_{b5}}}$$

where,

$AD^h_{it} =$ brand $i$'s share of advertising exposures received by household $h$ during 4 week period prior to occasion $t$,

$\pi^h_t =$ household $h$'s preference for brand $i$ on occasion $t$,

$REC^h_{it} =$ Recency of choice, 1 if $i$ was last brand purchased; $\alpha REC^h_{it}-1$ otherwise, where $\alpha$ is a smoothing constant,

$FEAT_{it} =$ 1 if brand $i$ is featured; 0 otherwise and

$DISP_{it} =$ 1 if brand $i$ is displayed; 0 otherwise.

The effect of achieving a certain share-of-voice is captured by the $(\gamma_{b0} + \gamma_{b1} \ln(\pi^h_t)) > 0$. Since $0 < \pi^h_t < 1$, a positive value for $\gamma_{b1}$ implies that a low-preference brand will have a smaller exponent value than a high-preference brand. Therefore, since $AD^h_{it}$ is a share measure, the same level of advertising will have a greater impact on inclusion probabilities for a low-preference brand than a high-preference one. In other words, a positive estimate of $\gamma_{b1}$ implies a negative interaction between brand advertising and preference.

Inferring brand preference: We infer brand preference from the choice history of households by assuming that the prior distribution of brand preference across households can be described by the Dirichlet distribution. Maximum likelihood estimates of the Dirichlet parameters are obtained from purchases observed in an initialization period. Knowing

One generic limitation of estimating the MCI model is that a variable should not take the value of 0. One way of circumventing this problem, as suggested by Cooper and Nakanishi (1988), is adding a small number (0.001) to the variable across brands and then rescaling the new value by the sum of the new values across brands. We apply this correction to our advertising share measure. For display and feature variables, we use 0.999 and 0.001 instead of 1 and 0. We also used the zeta-scores (Cooper and Nakanishi (1988)) and found similar results.
these parameters allows us to obtain a household's dynamically updated posterior brand preference ($\pi_{it}^h$) in a Bayesian fashion (See Siddarth et al. 1995 for details), incorporating both sample-wide priors (represented by the Dirichlet parameters) and household-specific purchase history (represented by total number of purchases of each brand from the initialization period up to the last purchase occasion).

### 3.4.4 Models Estimated

We estimate the proposed model along with several competing models described below.

**MNL Model**

This is the "plain" multi-segment logit choice model that accounts for preference and response heterogeneity but not for consideration sets, and assumes that consumers evaluate available brands in a compensatory fashion. The deterministic component of the MNL utility function for household $h$ in segment $s$ is given by:

$$U_{sit}^h = \beta_{s1} + \beta_{s2} AD_{it}^h + \beta_{s3} AD_{it}^h \ast \pi_{it}^h + \beta_{s4} \pi_{it}^h + \beta_{s5} REC_{it}^h + \beta_{s6} FEAT_{it} + \beta_{s7} DISP_{it},$$

(3.14)

where all variables are as defined above in Equation (3.13).

**Dynamic Bayes (DB) Model**

For comparison purposes, we also estimate the Dynamic Bayes (DB) model proposed by Siddarth et al. (1995). The DB model represents the restricted set approach to determining consideration sets, and employs a "crisp" version of a conjunctive rule. Andrews and Manrai (1995) suggest that the DB model performs better than previous consideration set models that have appeared in the marketing literature. The DB model specifies
brand choice probability as a mixture of two component probabilities as follows:

\[ P_{st}(i) = \lambda pc_{st}^h + (1 - \lambda) pf_{st}^h, \tag{3.15} \]

where \( pf_{st}^h \) is the probability of choosing brand \( i \) from all available brands, and \( pc_{st}^h \) is the probability of choosing brand \( i \) from the restricted choice set. The restricted choice set is determined by comparing dynamically updated brand preference (\( \pi_{it}^h \)) against a cross-sectional threshold level. The parameter \( \lambda \) can be interpreted as the probability that a household chooses from the restricted choice set.

**B&V Model**

As an alternative fuzzy set approach to modeling consideration sets, we also estimate a model based on B&V (1996). The deterministic component of the utility function is given by Equation (3.12) and is the same as the one used in our proposed model. However, the inclusion probability for each brand is given by a binary logit expression based on the following deterministic component of the consideration set utility function (termed the "salience function" in the B&V model):

\[ S_{st}^h = \gamma_{s0}AD_{st}^h + \gamma_{s1}AD_{st}^h \times \pi_{st}^h + \gamma_{s2} \pi_{st}^h + \gamma_{s3}REC_{st}^h + \gamma_{s4}FEAT_{st} + \gamma_{s5}DISP_{st}. \tag{3.16} \]

Then, \( IP_{st}^h \) is given by:

\[ IP_{st}^h = \frac{\exp(S_{st}^h)}{\exp(S_{st}^h) + \exp(\Delta)}, \tag{3.17} \]

where \( \Delta \) is a common threshold across brands, purchase occasions, and households, which is to be estimated. The more a brand’s salience exceeds \( \Delta \), the more the brand is likely to be included in the consideration set.
Proposed Models

We label the plain consideration set model as FZSET and the consideration set model with an entropy measure (price parameter as a function of relative entropy) as FZENT. The FZENT model subsumes the FZSET model and permits us to investigate the indirect effect of advertising on price sensitivity.

**FZSET Model:** We specify the overall utility function as follows:

\[
OU_{st}^h = \beta_0 + \beta_s \text{PRICE}_{st} + \gamma_{s0} \ln(\text{AD}^h_{st}) + \gamma_{s1} \ln(\text{AD}^h_{st}) \cdot \ln(\pi^h_{st}) + \gamma_{s2} \ln(\pi^h_{st}) + \gamma_{s3} \ln(\text{REC}^h_{st}) + \gamma_{s4} \ln(\text{FEAT}^h_{st}) + \gamma_{s5} \ln(\text{DISP}^h_{st}). \tag{3.18}
\]

As previously discussed, variables that affect consideration enter the overall utility function after being log-transformed. The final choice probability is then given by the expression, \(\frac{\exp(OU_{st}^h)}{\sum_k \exp(OU_{st}^k)}\) which is equivalent to Equation (3.8).

**FZENT Model:** We reparameterize \(\beta_{s1}\), the price response parameter in Equation (3.18) as a function of relative entropy as follows:

\[
\beta_{s1} = \theta_{s0} + \theta_1 \text{ENT}_{st}^h, \tag{3.19}
\]

where \(\text{ENT}_{st}^h\) is defined in Equation (3.11) and is segment-specific. This varying parameter specification enables us to study the relationship between consideration set size and price sensitivity after accounting for inherent differences in price sensitivity. \(\theta_{s0}\) is a segment-specific price parameter that reflects inherent (time-invariant) price sensitivity and \(\theta_1\) captures the impact of consideration set size on inter-temporal variations in price sensitivity.

We expect \(\theta_1\) to be negative, i.e., consideration set size to be positively related to price sensitivity. If this hypothesis is supported, it would imply that advertising affects inter-temporal price sensitivity through its influence on consideration set size.
Chapter 3. Consideration Set Heterogeneity, Advertising, and Price Sensitivity

3.5 Results and Discussion

For all the models, the two-segment solution provided the best overall fit based on the Bayesian Information Criteria (BIC). Fit statistics for this version of the models are provided in Table 3.1 (Table 3.2) for the yogurt (ketchup) category. Based on the BIC, the results show that all of the consideration set models outperform the MNL model (except for the B&V model in the ketchup category) and strongly support the notion that consumers simplify their choice decision by restricting consideration to a subset of available brands. Most importantly, the FZSET model outperforms the other consideration set models in both categories.

The superiority of the FZSET model over the DB model can be attributed to its utilization of continuous brand preference information to directly model inclusion probabilities as opposed to the discretized brand preference information used by the DB model. It is interesting to find that both the FZSET and DB models provide better fit than the B&V model (an exception is the holdout likelihood for the B&V model in the yogurt category). This supports the notion that consideration set formation is a non-compensatory process as described by the FZSET (a fuzzy version) and DB (a crisp version) models, rather than a compensatory process as represented by the B&V model (Note that, in Equation (3.10), the B&V model appears to be algebraically equivalent\textsuperscript{12} to the MNL model (compensatory process) except for the negative log term.) Based on these results, it seems that the FZSET model provides a better description of the choice processes than do other models.

Parameter estimates for the models are reported in Tables 3.3 and 3.4. In the yogurt category (Table 3.3), price parameters are all significant and correctly signed. Feature advertising is not estimated to have a significant effect in any of the models.\textsuperscript{13}

\textsuperscript{12}In the ketchup category, the B&V model turns out to be almost equivalent to the MNL model

\textsuperscript{13}Only two brands carried feature ads and the frequency is extremely low (58 times out of 3053
The FZSET model is the only one in which, both, the main and interaction effects of advertising are significant in Segment 1. As previously discussed, a positive value for $\gamma_1$ in the FZSET model indicates that the effect of advertising on inclusion probabilities depends on brand preferences. More specifically, advertising’s impact is found to be higher for less preferred brands. In contrast, the B&V model results show a significant positive interaction effect (but insignificant main effect of advertising) in Segment 1. Given the superior fit of the FZSET model in both the calibration and holdout samples, we argue that, for the yogurt category, the impact of advertising on inclusion probabilities is likely to be higher for less preferred brands than for more preferred brands. This effect holds for Segment 1 that accounts for more than 60% of the sample households, but not for Segment 2.

In the ketchup category (Table 3.4), we do not investigate the effect of advertising due to data limitations. All estimated parameters are correctly signed and reasonable in terms of face validity.

It is noteworthy that the smoothing parameter $\alpha$ in the FZSET model is substantially lower than those in the comparison models, indicating that consumers give more weight to the brand last purchased than the comparison models predict. Due to differences in model formulations, the magnitude of variables that affect consideration cannot be directly compared (In the next section, we report an analysis of elasticities for a direct comparison).

Fit statistics and parameter estimates for the FZENT model, estimated on the yogurt data, are presented in Table 3.5. The results show that the FZENT and FZSET models are basically equivalent, and do not provide the evidence for a positive relationship between consideration set size and price sensitivity. In contrast to the experimental findings of Mitra and Lynch (1995), the hypothesized effects do not appear to be carried over to purchase occasions).
this particular data set. Unlike their experimental study in which unfamiliar brands are used, we study a well-established market (yogurt) which consists of familiar brands (no new brands). In this market, advertising may simply reinforce brand preference rather than change consideration set size, resulting in a smaller variation in the set size than we expect in the market that consists of all unfamiliar brands.

3.6 Managerial Implications

Dickson and Sawyer (1990) have shown that consumers spend only a limited amount of time at the point of purchase and that they may not examine all the prices of available brands in the category. In other words, consumers evaluate the prices of a subset of brands. In contrast to this empirical finding, previous single-stage choice models have assumed that consumers evaluate the prices of all available brands. A well-known consequence of ignoring consideration set formation in modeling brand choice is that the impact of price may be biased as compared to the "true" impact which would be obtained if the model was specified only for actively considered alternatives (Meyer and Kahn 1991). Since consideration is a necessary condition for choice, it is critical to identify a brand's position in the consideration set in order to decide on relative emphasis on marketing communication elements that affect consideration set formation vs. pricing. For example, if we can identify brands that have low inclusion probabilities, then, in the presence of price discounts across brands, such brands can benefit more by increasing their communication efforts since price cuts for brands not likely to be considered may now become noticed.

In Table 3.6, we present the average inclusion probabilities (based on the FZSET model) and market shares for the calibration sample. As shown in Siddarth et al. (1995),
the rank-order of brands by market share may not be the same as the rank-order by inclusion probability in the current study. For example, Yoplait and Dannon rank second and fourth, respectively, in inclusion probability but rank third and fifth in market share. On the other hand, W.B.B and Weight Watcher rank third and fifth in inclusion probability but rank second and fourth in market share. This suggests a potentially different focus in target consumers and marketing effort across brands. Yoplait and Dannon might want to induce more purchases among current customers by providing financial incentives while W.B.B and Weight Watcher might want to attract new customers by focusing on marketing communications.

In principle, any brand needs a base-line level of marketing communications but some brands might need more so that their pricing may have an impact on market share. To illustrate this point, we conducted a simulation based on the parameter estimates obtained from the FZSET model. First, we simulated a 10% price cut and calculated market shares given the current level of advertising and display. Then, with the same price cut, we simulated one additional TV ad exposure\(^\text{14}\) prior to each purchase occasion and display in 50 randomly selected week-stores (out of 663 week-stores), respectively, and calculated market shares again. Simulation results are shown in Table 3.7.

Results show that additional TV ads and displays, respectively, appear to amplify the impact of price cut on share gains and that the share gains vary across brands due to differences in inclusion probabilities. The impact of additional displays, in particular, on share gains is greater especially for the two brands that have the lowest inclusion probabilities, QC and private label. This suggests that brands, which have low inclusion probabilities need some type of communication, e.g., displays, otherwise their price promotion may go unnoticed.

\(^{14}\)Four brands are excluded from this simulation because they never or rarely advertised on TV during the study period and therefore, the condition given here is too different from what actually happened, to make a reliable prediction based on the estimated parameters.
Chapter 3. Consideration Set Heterogeneity, Advertising, and Price Sensitivity

We repeated a simulation based on the parameter estimates obtained from the MNL model. Simulation results are shown in Table 3.8. As compared to the share gains predicted by the FZSET model, the MNL model appears to underpredict share gains from improving pricing and communication mix elements, e.g., displays. This is the expected result since the MNL model dilutes the effect of the communication mix, assuming that all brands are considered, whereas the FZSET Model sharpens the effect of the communication mix, assuming that consideration is a necessary condition for choice. Therefore, the simulation results illustrate the importance of incorporating the consideration set into brand choice model to capture accurately the impact of a price discount coupled with improved communication mix.

3.7 Conclusion

We develop a new fuzzy consideration set model that incorporates the effects of brand preference, recency of choice, TV ads, feature ads, and display on consideration set formation. As opposed to the universal set approach or the restricted set approach that model uncertainty about the sets from which a brand is chosen, the proposed fuzzy set approach directly models the probability of a brand being included in the consideration set. The resulting inclusion probabilities for brands reflect a "fuzzy" consideration set, in the sense that a brand belongs to the consideration set only probabilistically, in contrast to well-defined consideration sets, in which consumers are assumed to know which brands are included in each consideration set but there is uncertainty about the sets from which a brand will be chosen.

As compared to an existing fuzzy set model (B&V 1996) in which both brand evaluation and consideration variables have a compensatory influence, our model permits a

\footnote{Additional ad exposures were not simulated since the advertising parameter in the MNL model is insignificant.}
non-compensatory process in consideration set formation while allowing brand evaluation to be compensatory.

Results from estimating our model on scanner panel data from the yogurt and ketchup categories shows that the FZSET model outperforms the DB model. The superiority of the FZSET model can be attributed to utilization of continuous brand preference information as opposed to the discretized information used in the DB model. In addition, both the FZSET and DB models provide a better fit than the B&V model, implying that consideration set formation is a non-compensatory, rather than compensatory, process.

To investigate the indirect effect of advertising on price sensitivity through its influence on the consideration set, we extend the FZSET model by reparameterizing the price parameter as a function of relative entropy (FZENT model) which measures consideration set size. We find no evidence for a positive relationship between consideration set size and price sensitivity. The experimental results of Mitra and Lynch (1995) do not seem to be carried over to this particular data set. It may be harder to detect variation in consideration set size in the established market (yogurt) that consists of relatively familiar brands (no new brands). Further empirical test may be necessary to confirm their hypothesized relationship with other data set that includes both existing and new brands.

Simulation results based on estimated parameters show that the impact of TV ads and display on share gains vary across brands due to differences in inclusion probabilities. This suggests that marketing communications, e.g., display, can amplify the impact of a price cut on share gains especially for brands that have low inclusion probabilities. As compared to the share gains predicted by our model, the MNL model appears to under-predict share gains by additional display, that is, underestimate the impact of marketing communications. This illustrate the importance of incorporating the consideration set into a brand choice model to assess accurately the impact of marketing communications.
3.8 Limitations and Future Research

In investigating the effect of advertising on price sensitivity through its influence on consideration set size, we used a share-of-voice measure without taking into account advertising content, i.e., price-oriented vs. image-oriented advertising. As speculated in Boulding et al. (1994), the content of advertising could influence price sensitivity qualitatively, i.e., a share of voice and advertising content may jointly affect price sensitivity through their impact on the consideration set size. In addition, we studied a well-established market (yogurt) which consisted of relatively familiar brands (no new brands) unlike the experimental study of Mitra and Lynch (1995). Therefore, it may have been harder to detect variation in consideration set size and thus a relationship between consideration set size and price sensitivity. Further investigation may be necessary to generalize the findings of Mitra and Lynch (1995) with other data set that contains more accurate and richer information about advertising for both existing and new brands.

We specify the overall utility function that incorporates two sets of variables that might affect consideration set formation and brand evaluation, respectively, based on the classification of previous research (Nedungadi 1990, Andrews and Srinivasan 1995, Siddarth et al. 1995, B&V 1996). We assume that consumers uniformly screen out brands based on these pre-specified variables. However, consumers may differ not only in their consideration sets but also in their choice of variables on which they screen out brands. Therefore, from the analyst’s point of view, there is uncertainty about which variables consumers would use in consideration set formation. We could incorporate this uncertainty into a more general model that allows for the set of variables that influence the consideration set to vary across consumers.

One additional issue in research on consideration set formation is whether one brand
can belong to a consideration set without affecting the probability of another brand belonging to the consideration set. The proposed fuzzy set model assumes “dependence” among inclusion probabilities, that is, increased inclusion probability for one brand necessarily decreases the sum of inclusion probabilities for remaining brands. However, it can be argued any brand can belong to a consideration set independently (Andrew and Srinivasan 1995, B&V 1996). A further generalization may be that one segment of consumers may follow the “dependence” assumption and another segment of consumers may follow the “independence assumption” based on their cognitive capacity or willingness to carry brands in the consideration set. Therefore, we could incorporate these competing views into an integrated model to describe better the consideration set formations across segments. Essentially, this generalization seek to account for model uncertainty by allowing structural assumptions to vary across segments. We can deal with this issue by employing a Bayesian model averaging approach, as discussed in the next chapter.
Table 3.1: Fit Statistics for the Yogurt Data

<table>
<thead>
<tr>
<th>Model</th>
<th>Number of Parameters</th>
<th>Calibration log-likelihood</th>
<th>BIC</th>
<th>Holdout log-likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full Set</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MNL</td>
<td>22</td>
<td>-3137.531</td>
<td>-3225.794</td>
<td>-1424.837</td>
</tr>
<tr>
<td>Consideration Set</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DBM</td>
<td>24</td>
<td>-3121.569</td>
<td>-3217.856</td>
<td>-1419.600</td>
</tr>
<tr>
<td>B&amp;V</td>
<td>23</td>
<td>-3131.227</td>
<td>-3223.502</td>
<td>-1418.532</td>
</tr>
<tr>
<td>FZSET</td>
<td>22</td>
<td>-3098.374</td>
<td>-3186.637</td>
<td>-1383.128</td>
</tr>
</tbody>
</table>

2 segment solution is the best based on BIC for all models.
### Table 3.2: Fit Statistics for the Ketchup Data

<table>
<thead>
<tr>
<th>Model</th>
<th>Number of Parameters</th>
<th>Calibration log-likelihood</th>
<th>BIC</th>
<th>Holdout log-likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Full Set</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MNL</td>
<td>18</td>
<td>-2701.921</td>
<td>-2772.251</td>
<td>-1321.615</td>
</tr>
<tr>
<td><strong>Consideration Set</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DBM</td>
<td>20</td>
<td>-2674.621</td>
<td>-2752.765</td>
<td>-1306.080</td>
</tr>
<tr>
<td>B&amp;V</td>
<td>19</td>
<td>-2701.921</td>
<td>-2776.158</td>
<td>-1321.617</td>
</tr>
<tr>
<td>FZSET</td>
<td>18</td>
<td>-2665.829</td>
<td>-2736.159</td>
<td>-1298.054</td>
</tr>
</tbody>
</table>

2 segment solution is the best based on BIC for all models.


Chapter 3. Consideration Set Heterogeneity, Advertising, and Price Sensitivity

Table 3.3: Parameter Estimates for the Yogurt Data

<table>
<thead>
<tr>
<th>Variable</th>
<th>MNL</th>
<th>DBM</th>
<th>B&amp;V</th>
<th>FZSET</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SEG1</td>
<td>SEG2</td>
<td>SEG1</td>
<td>SEG2</td>
</tr>
<tr>
<td>PRICE</td>
<td>-0.694</td>
<td>-0.307</td>
<td>-0.695</td>
<td>-0.323</td>
</tr>
<tr>
<td>AD</td>
<td>0.018*</td>
<td>0.519*</td>
<td>0.076*</td>
<td>0.573*</td>
</tr>
<tr>
<td>AD*π</td>
<td>1.774*</td>
<td>-0.706*</td>
<td>1.378*</td>
<td>-0.794*</td>
</tr>
<tr>
<td>π</td>
<td>1.190</td>
<td>2.232</td>
<td>1.023</td>
<td>2.011</td>
</tr>
<tr>
<td>REC</td>
<td>0.708</td>
<td>2.043</td>
<td>0.636</td>
<td>2.042</td>
</tr>
<tr>
<td>FEAT</td>
<td>0.318*</td>
<td>-0.306*</td>
<td>0.398*</td>
<td>-0.417*</td>
</tr>
<tr>
<td>DISP</td>
<td>1.584</td>
<td>0.686</td>
<td>1.644</td>
<td>0.653</td>
</tr>
<tr>
<td>Smoothing α</td>
<td>0.836</td>
<td>0.823</td>
<td>0.790</td>
<td>0.291</td>
</tr>
<tr>
<td>Cutoff (δ)</td>
<td>0.052</td>
<td>0.260</td>
<td>3.963</td>
<td></td>
</tr>
<tr>
<td>λ</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Threshold (Δ)</td>
<td></td>
<td></td>
<td></td>
<td>3.963</td>
</tr>
<tr>
<td>Segment Share</td>
<td>0.640</td>
<td>0.360</td>
<td>0.681</td>
<td>0.319</td>
</tr>
</tbody>
</table>

(*): insignificant at α = 0.05; All other parameters are significant at α = 0.05
### Table 3.4: Parameter Estimates for the Ketchup Data

<table>
<thead>
<tr>
<th>Variable</th>
<th>MNL</th>
<th>DBM</th>
<th>B&amp;V</th>
<th>FZSET</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SEG1</td>
<td>SEG2</td>
<td>SEG1</td>
<td>SEG2</td>
</tr>
<tr>
<td>PRICE</td>
<td>-1.667</td>
<td>-0.121*</td>
<td>-1.623</td>
<td>-0.130*</td>
</tr>
<tr>
<td>( \pi )</td>
<td>4.270</td>
<td>3.773</td>
<td>3.798</td>
<td>2.838</td>
</tr>
<tr>
<td>REC</td>
<td>0.607</td>
<td>0.741</td>
<td>0.415</td>
<td>0.752</td>
</tr>
<tr>
<td>FEAT</td>
<td>0.618</td>
<td>0.808*</td>
<td>0.675</td>
<td>0.828*</td>
</tr>
<tr>
<td>DISP</td>
<td>2.075</td>
<td>2.592</td>
<td>2.210</td>
<td>2.837</td>
</tr>
<tr>
<td>Smoothing ( \alpha )</td>
<td>0.961</td>
<td>0.892</td>
<td>0.961</td>
<td></td>
</tr>
<tr>
<td>Cutoff (( \delta ))</td>
<td></td>
<td>0.070</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \lambda )</td>
<td></td>
<td>0.408</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Threshold (( \Delta ))</td>
<td></td>
<td></td>
<td></td>
<td>16.673*</td>
</tr>
<tr>
<td>Segment Share</td>
<td>0.717</td>
<td>0.283</td>
<td>0.749</td>
<td>0.251</td>
</tr>
</tbody>
</table>

(*): insignificant at \( \alpha = 0.05 \); All other parameters are significant at \( \alpha = 0.05 \)
Table 3.5: FZENT: Yogurt Data

<table>
<thead>
<tr>
<th>Variable</th>
<th>SEG1</th>
<th>SEG2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Brand Evaluation</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PRICE ((\theta_{s0}))</td>
<td>-0.734</td>
<td>-0.333</td>
</tr>
<tr>
<td>ENT*PRICE ((\theta_1))</td>
<td>0.066*</td>
<td></td>
</tr>
<tr>
<td><strong>Consideration Set Formation</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AD ((\gamma_{s0}))</td>
<td>0.063</td>
<td>0.027*</td>
</tr>
<tr>
<td>AD*π ((\gamma_{s1}))</td>
<td>0.014</td>
<td>-0.001*</td>
</tr>
<tr>
<td>π ((\gamma_{s2}))</td>
<td>0.376</td>
<td>0.425</td>
</tr>
<tr>
<td>REC ((\gamma_{s3}))</td>
<td>0.043</td>
<td>0.116</td>
</tr>
<tr>
<td>FEAT ((\gamma_{s4}))</td>
<td>0.036*</td>
<td>-0.044*</td>
</tr>
<tr>
<td>DISP ((\gamma_{s5}))</td>
<td>0.232</td>
<td>0.087</td>
</tr>
<tr>
<td>Smoothing (α)</td>
<td>0.294</td>
<td></td>
</tr>
<tr>
<td><strong>Fit Statistics</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Calibration Log-likelihood</td>
<td>-3098.186</td>
<td></td>
</tr>
<tr>
<td>BIC</td>
<td>-3190.460</td>
<td></td>
</tr>
<tr>
<td>Holdout Log-likelihood</td>
<td>-1384.302</td>
<td></td>
</tr>
</tbody>
</table>

\(^*\): insignificant at \(α = 0.05\)
### Table 3.6: Inclusion Probability and Market Share

<table>
<thead>
<tr>
<th>Brand</th>
<th>Inclusion Probability</th>
<th>Market Share</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yoplait</td>
<td>0.172</td>
<td>0.160</td>
</tr>
<tr>
<td>Weight Watcher</td>
<td>0.114</td>
<td>0.135</td>
</tr>
<tr>
<td>Dannon</td>
<td>0.144</td>
<td>0.122</td>
</tr>
<tr>
<td>Nordica</td>
<td>0.305</td>
<td>0.268</td>
</tr>
<tr>
<td>Quality Control (QC)</td>
<td>0.024</td>
<td>0.035</td>
</tr>
<tr>
<td>Well's Blue Bunny (W.B.B)</td>
<td>0.147</td>
<td>0.175</td>
</tr>
<tr>
<td>Private Label</td>
<td>0.094</td>
<td>0.105</td>
</tr>
</tbody>
</table>
Table 3.7: Share Gains Across Brands: FZSET

<table>
<thead>
<tr>
<th>Brand</th>
<th>10% Price Cut with Current Ad Exposures and Display Level</th>
<th>10% Price Cut with Additional Ad Exposures</th>
<th>10% Price Cut with Additional Displays</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yoplait</td>
<td>4.33%</td>
<td>4.91%</td>
<td>4.68%</td>
</tr>
<tr>
<td>Weight Watcher</td>
<td>2.34%</td>
<td>2.41%</td>
<td>2.59%</td>
</tr>
<tr>
<td>Dannon</td>
<td>3.08%</td>
<td>3.44%</td>
<td>3.41%</td>
</tr>
<tr>
<td>Nordica</td>
<td>3.81%</td>
<td>**</td>
<td>4.44%</td>
</tr>
<tr>
<td>Quality Control</td>
<td>2.34%</td>
<td>**</td>
<td>8.24%</td>
</tr>
<tr>
<td>Well's Blue Bunny</td>
<td>2.75%</td>
<td>**</td>
<td>3.37%</td>
</tr>
<tr>
<td>Private Label</td>
<td>2.92%</td>
<td>**</td>
<td>7.73%</td>
</tr>
</tbody>
</table>

**: These brands are excluded from the simulation.
Table 3.8: Share Gains Across Brands: MNL

<table>
<thead>
<tr>
<th>Brand</th>
<th>10% Price Cut with Current Display Level</th>
<th>10% Price Cut with Additional Displays</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yoplait</td>
<td>3.95%</td>
<td>4.17%</td>
</tr>
<tr>
<td>Weight Watcher</td>
<td>2.60%</td>
<td>2.84%</td>
</tr>
<tr>
<td>Dannon</td>
<td>3.01%</td>
<td>3.23%</td>
</tr>
<tr>
<td>Nordica</td>
<td>3.23%</td>
<td>3.65%</td>
</tr>
<tr>
<td>Quality Control</td>
<td>2.09%</td>
<td>6.85%</td>
</tr>
<tr>
<td>Well's Blue Bunny</td>
<td>2.42%</td>
<td>2.86%</td>
</tr>
<tr>
<td>Private Label</td>
<td>2.50%</td>
<td>6.23%</td>
</tr>
</tbody>
</table>
Figure 3.1: **Functional Shapes of Compensatory vs. Non-Compensatory Rule**

* Compensatory: $U = \gamma y$ where $\gamma$ is set equal to 0.3 and $y$ represents one attribute

* Fuzzy non-compensatory: $U = \gamma \ln(y)$ where $\gamma$ is set equal to 0.1

* Crisp non-compensatory: $U = -\infty$ for $y \leq \delta$, otherwise 0 where $\delta$ is a threshold and set equal to 0.05 (cut off value of the Dynamic Bayes Model (Siddarth et al. 1995) estimated in the current study)
Figure 3.2: Relationship Between Advertising and Price Sensitivity
4.1 Overview

Faced with competing hypotheses, most often, we develop different models based on their respective structural assumptions, examine the data to identify a single best model, and then proceed as if the chosen model based on fit statistics were known to be true. The chosen model is then used to make inferences and predictions even though the rejected models are not necessarily wrong (Draper 1995). In other words, structural uncertainty about the model itself is generally neglected.

In the statistics literature, the issue of model uncertainty (or structural uncertainty) has been addressed by the Bayesian model averaging approach (Chatfield 1995) that acknowledges the possibility that there may be more than one model which may be regarded as "close to true". The notion of having more than one model is a key element of the Bayesian model averaging approach which avoids having to select a single best model but rather averages over more than one model. The spirit of this approach is well represented by the following quote:

"The main way to avoid noticing after the fact that a set of modeling assumptions, different from those originally assumed, turned out to be correct is for one's model prospectively to have been sufficiently large to encompass the retrospective truth (Draper 1995, p. 55)."

Recently, two studies that account for structural or model uncertainty were introduced
in the marketing literature. Kamakura, Kim and Lee (1996) develop a flexible model in which consumers may differ in the choice process (i.e., use either a brand-primary or a form-primary decision rule). In an analysis based on store-level data, Kalyanam (1996) applies a Bayesian mixture model that accounts for uncertainty about functional forms of demand, recognizing that in practice many different demand functions might be theoretically plausible and fit the data equally well.

This type of modeling approach, that incorporates competing yet plausible hypotheses in an integrated model, seeks to build a “big” model, which encompasses all plausible hypotheses, rather than using a single best model, which may not represent household-level choice behavior uniformly better than other, rejected, models. In deciding how big is sufficiently big, we may resort to subject-matter considerations such as accepted theories, expert background knowledge, and prior information, including that obtained from previous, similar, data sets.

In marketing, a significant body of research has investigated the reference price construct and modeled its impact on consumer choice behavior. Two major competing models have been proposed: the sticker shock (Winer 1986) and reference-dependent formulations (Hardie, Johnson and Fader 1993). The sticker shock formulation captures reference price effects via “a sticker shock term” defined as the difference between brand-specific reference prices and the current prices. On the other hand, the reference-dependent formulation captures asymmetric responses to positive and negative price deviations from a category-specific reference price, which are hypothesized to result in loss aversion (Tversky and Kahneman 1991) in the case of the current price exceeding the reference price.

---

\[\text{We may allow for different parameters for the positive and negative differences to capture asymmetric responses to price losses and gains (e.g., Kalwani et al. 1990, Briesch et al. 1997).}\]
Chapter 4. Decision Rule Heterogeneity

A key distinction between the two models is that the sticker shock formulation assumes that consumers compare the current price of each alternative to its brand-specific reference price while the reference-dependent formulation assumes that consumers compare the current price of each alternative to one, category-specific, reference price and that consumer responses to gains and losses are asymmetric. Therefore, these two models represent two different decision rules or choice processes underlying consumer brand choice, both of which are empirically supported and theoretically grounded in psychological theory: Adaptation-Level Theory (Helson 1964) and Prospect Theory (Kahneman and Tversky 1979). Then, naturally arising questions are, "Which model represents the true nature of response to reference price" and "Which model should we choose"?

In an attempt to answer these questions, Briesch et al. (1997) performed a comparative analysis of brand-specific and category-specific reference price models. Their empirical results from four product categories provided overall support for the memory-based brand-specific reference price model (i.e., the sticker shock model). They assume that all consumers consistently follow either the sticker shock model or reference-dependent model, and once a single best model is chosen, they base their subsequent inferences on that model. Is this view correct? More importantly, if one model is better supported by data, does that model represent the whole truth? Is it possible that the other, rejected, models may also partially represent reality? These questions are all related to the issue of model uncertainty.

In this chapter, we develop a model that accounts for uncertainty about the way reference prices affect brand choice. We avoid selecting a single best model but, rather,

They evaluated five alternative models of reference price which are either stimulus-based (i.e., based on information available at the point-of-purchase) and memory-based (i.e., based on price history). Essentially, however, these models can be regrouped into the sticker shock and reference-dependent models, depending on whether reference price is brand-specific or category-specific.
incorporate the competing decision rules into a single model to account for model uncertainty. Our approach recognizes the possibility that there may be two consumer segments, one consistently following the sticker shock model while the other consistently follows the reference-dependent model. In addition, as addressed in Briesch et al. (1997), consumers may use either of the decision rules on different purchase occasions, i.e., mix the two decision rules\(^3\) probabilistically. In summary, we develop a general (big) model that incorporates the distinct model structures implied by the three different decision rules.

This research seeks to provide the following contributions:

- Develop a structural uncertainty-based model of reference price effects that allows for different decision rules (decision rule heterogeneity).

- Demonstrate, by way of an analysis of scanner panel data, that properly accounting for structural uncertainty can provide a better description of the way reference prices affect the choice process.

This chapter is organized as follows. In Section 4.2, we present a model that fully accounts for both structural uncertainty and parametric uncertainty (parameter heterogeneity across consumers) in capturing reference price effects. In Section 4.3 we compare our approach to one that ignores model uncertainty. Section 4.4 describes sampling-based methods (Markov Chain Monte Carlo simulation) to estimate the proposed model. Section 4.5 presents an empirical application to scanner panel data from the ketchup category. Section 4.6 discusses the empirical results. We close with our conclusions and directions for future research.

\(^3\)In fact, this mixed rule could have been incorporated into Kamakura et al's study (1996) since \textit{a priori} it is uncertain whether consumers consistently make either brand-primary decision or product form-primary decision, or they simply mix the two decision rules probabilistically on each purchase occasion.
4.2 Model Development

We specify three models (structures) that consumers follow in their brand choice decision. A consumer could be described by a sticker shock model \((M = 1)\), a reference-dependent model \((M = 2)\), or a mix of the two decision rules \((M = 3)\).

Given a particular state \(M = s\), we specify the probability of brand choice for household \(h\) on occasion \(t\) as the Multinomial Logit (MNL) probability:

\[
P^h_{st}(i \mid \beta^h_s) = \frac{\exp(U^h_{sit})}{\sum_k \exp(U^h_{skt})},
\]

where \(U^h_{sit}\) denotes the deterministic component of utility for alternative \(i\), household \(h\), and model structure \(s\), and \(\beta^h_s\) is a household- and structure-specific parameter vector.

**Structure 1**: The sticker shock model reflects a decision rule that incorporates responses to the differences between brand-specific reference prices and shelf prices. Following Winer(1986) and Lattin and Bucklin(1989), the sticker shock utility\(^4\) is given by:

\[
U^h_{1it} = \beta^h_{10i} + \beta^h_{11} \text{PRICE}_{it} + \beta^h_{12} (\text{RPRICE}^h_{it} - \text{PRICE}_{it}) + \beta^h_{13} \text{PROMO}_{it},
\]

where \(\text{RPRICE}^h_{it}\) is a brand-specific reference price. The reference price of a brand is assumed to be its price\(^5\) on the last occasion on which the household made a purchase.

**Structure 2**: The reference-dependent model assumes that consumers compare the current price of each alternative to one, category-specific, reference price and that consumer responses to gains and losses are asymmetric. Following Hardie, Johnson and Fader (1993), the reference-dependent utility is given by:

\[
U^h_{2it} = \beta^h_{20i} + \beta^h_{21} \text{PGAIN}_{it} + \beta^h_{22} \text{PLOSS}_{it} + \beta^h_{23} \text{PROMO}_{it},
\]

\(^4\)In the empirical application, we include two additional variables, brand loyalty and last brand purchased, that are often used in a standard logit model.

\(^5\)Alternatively, Briesch et al. (1997) used the exponentially smoothed (memory-based) shelf prices observed on purchase occasions. In our application, we choose to use the simplest form of the reference price since we are interested in structural uncertainty rather than in alternative operationalizations.
where $\text{PGAIN}^h_{i_t}$ is the difference between the reference price and actual price of brand $i$ when the actual price is lower than the reference price, 0 otherwise and $\text{PLOSS}^h_{i_t}$ is the difference between the reference price and actual price of brand $i$ when the actual price is higher than the reference price, 0 otherwise. The reference price of a brand is assumed to be the price last paid\(^6\) for a purchase in the category.

**Structure 3:** The mixed decision rule is represented by the following model:

$$P^h_{3t}(i | \beta^h_3) = \gamma^h P^h_{1t}(i | \beta^h_{31}) + (1 - \gamma^h) P^h_{2t}(i | \beta^h_{32}) \quad (4.4)$$

where $\gamma^h$ represent the probability that household $h$ follows the sticker shock model on each purchase occasion and $\beta^h_3 = (\beta^h_{31}, \beta^h_{32}, \gamma^h)$.

As compared to Structures 1 and 2, that represent either a pure sticker shock or a pure reference-dependent decision rule across all purchase occasions, Structure 3 implies a probabilistic mixture of the two decision rules, which is consistent over time.

### 4.2.1 Model Uncertainty

It is uncertain which decision rule a particular household $h$ may follow. To account for this uncertainty, we specify the household-specific probability that household $h$ follows decision rule $s$ as $p^h(M = s) = \lambda^h_s$, where $\sum_{s=1}^{S} \lambda^h_s = 1$.

**Priors on Model Probabilities:** If the model indicator $M$ for household $h$ is thought of as a multinomial random variable with probability vector $\lambda^h$, then the natural conjugate prior distribution is the Dirichlet distribution, i.e., $\lambda^h \sim D(\alpha_1, \alpha_2, \alpha_3)$. The relative sizes of the Dirichlet parameters $\alpha_s$ describe the mean of the prior distribution for $\lambda^h_s$, and

\(^6\)In Briesch et al. (1997), two alternative operationalizations of the reference price are suggested. One is the current (stimulus-based) shelf price of the reference brand (the chosen brand on the last purchase occasion). The other is the exponentially smoothed shelf prices of the brands chosen on past purchase occasions (memory-based). Following the standard approach in previous research, we choose to use the price last paid (memory-based) as the reference price. As compared to the exponentially smoothed shelf prices, this measure is simpler and avoids confounding prices of different brands and sizes.
Chapter 4. Decision Rule Heterogeneity

the sum of the \( \alpha_s \)'s is a measure of the strength of the prior distribution, i.e., the prior sample size. This specification accounts for structural uncertainty within a household and heterogeneity in decision rules across households.

4.2.2 Parameter Heterogeneity

Conditional on the model indicator \( M = s \), it is straightforward to model heterogeneity in parameters across households. Following the standard approach suggested in the literature (e.g., Gelfand et al. 1990), we specify a prior distribution over household parameters, \( \beta^h_s \), given \( M = s \).

*Priors over Household Parameters:* Since the fully Bayesian approach requires priors over unknown quantities in the model, we specify a multivariate normal prior distribution over household parameters, \( \beta^h_s \), given \( M = s \).

\[
\beta^h_s | M = s \sim MVN(\mu_s, \Sigma_s)
\]  

(4.5)

We can either estimate \( \mu_s \) and \( \Sigma_s \) or pre-specify their values by some constants.\(^7\) Since we are interested in population quantities, \( \mu_s \) and \( \Sigma_s \), we estimate them by specifying hyper-priors over them. In this stage, we fix the parameter values of the hyper-priors such that they represent diffuse but proper distribution. Following Gelfand et al. (1990), we assume a multivariate normal prior, \( MVN(\eta_s, C_s) \) over \( \mu_s \) and a Wishart prior, \( W[(\rho_s R_s)^{-1}, \rho_s] \) over \( \Sigma_s \). Details on each prior are provided in Appendix C.

4.2.3 Joint Model Specification

Given the model probabilities and priors over household parameters, we can specify the joint distribution of all unknown parameters and data, \( y^h \) (vector of indicators for chosen

\(^7\)Pre-specifying the parameter values for a prior distribution is a simplification of the fully Bayesian approach and is similar to the approach suggested in Rossi and Allenby (1993). This approach ignores uncertainty in the parameters of the prior.
brands). Corresponding to model structure $s$, let $p(\beta^h_s \mid M = s)$ be a prior over household parameters, and $f(y^h \mid \beta^h_s, M = s) = \prod_{t=1}^{T_h} P_{st}(c \mid \beta^h_s)$ be a likelihood for household $h$, where $c$ is an indicator for a chosen brand. Conditional on $M = s$, only $\beta^h_s$ is relevant to $\Pi_{t=1}^{T_h} P_{st}(c \mid \beta^h_s)$ (Similarly, only $p(\beta^h_s \mid M = s)$ is relevant to $\beta^h_s$). In other words, conditional on $M = s$, $\beta^h_s$ is assumed to describe completely $y^h$, the observed choice history of household $h$. Therefore, we assume complete independence among the various $\beta^h_s$ given $M = s$ (i.e., $y^h$ is independent of $\{\beta^h_{k \neq s}\}$ given $M = s$), and may specify the fully Bayesian model by choosing proper “pseudopriors”, $p(\beta^h_s \mid M \neq s)$ (We discuss the role of pseudoprior and the specifics of pseudoprior selection, as described in Carlin and Chib 1995, in Appendix C). From the usual conditional independence assumptions,

$$p(y^h \mid M = s) = \int f(y^h \mid \beta^h, M = s)p(\beta^h \mid M = s)d\beta^h$$

$$= \int f(y^h \mid \beta^h_s, M = s)p(\beta^h_s \mid M = s)d\beta^h_s,$$

where $\beta^h = (\beta^h_1, \beta^h_2, \beta^h_3)$, and

$$p(y^h) = \sum_{s=1}^{3} \lambda^h_s \int f(y^h \mid \beta^h_s, M = s)p(\beta^h_s \mid M = s)d\beta^h_s. \quad (4.6)$$

So the form given to $p(\beta^h_s \mid M \neq s)$ is irrelevant in Equation (4.6). As the name suggests, a pseudoprior is not really a prior but only a conveniently chosen linking density (Carlin and Chib 1995), required to define the following complete joint model specification. Given model probability $p^h(M = s) = \lambda^h_s$, the joint distribution of $y^h$ and $\beta^h$ when $M = s$ is given by:

$$p(y^h, \beta^h, M = s) = f(y^h \mid \beta^h_s, M = s) \left[ \prod_{k=1}^{3} p(\beta^h_k \mid M = s) \right] \lambda^h_s, \quad (4.7)$$

Unlike in Equation (4.6), pseudopriors, e.g., $p(\beta^h_1 \mid M = 1)$ and $p(\beta^h_3 \mid M = 1)$ need to be specified in Equation (4.7). Given this joint distribution, we account for structural and parametric uncertainty by integrating over uncertainty about model parameters, $\beta^h$ and $\lambda^h$, respectively.
Given the priors and likelihood, we construct the full conditional distributions for the entire model parameters and implement the MCMC simulations to estimate posterior distributions of model parameters. The exact forms of the full conditional distributions and detailed estimation procedures are discussed in Appendix C.

4.3 Model Features

Our modeling approach explicitly incorporates structural uncertainty into the model, recognizing that any possible structure receiving a prior probability of 0 must also have a posterior probability of 0. Given that both the sticker shock model and the reference-dependent model of reference price are theoretically grounded and empirically supported, it is imperative to put non-zero prior model probabilities on these structures as opposed to setting $p_h(M = s) = 1.0$ and $p_h(M \neq s) = 0.0$, i.e., putting a point mass on only one structure $s$, which is equivalent to estimating each model separately and choosing a single best model in terms of fit statistics. This "point-mass-on-one-structure-at-a-time" may be too concentrated on a single structural assumption to lead to well-calibrated inferences and predictions (Draper 1995).

Behaviorally, there is no reason to believe that every household follow the same decision rule characterized either by the sticker shock model or by the reference-dependent model. A consumer segment may follow the sticker shock model because they have sufficient cognitive capacity as well as motivation to recall the shelf-prices observed on the last purchase occasion. Another segment may follow the reference-dependent model because they want to simplify the decision process and reduce their cognitive burden by having only one, category-specific, reference price. The other segment may mix the two decision rules probabilistically. For a particular household $h$, we can infer which decision rule the household may follow by examining the model probability, $p_h(M = s) = \lambda^h_s$. 
Chapter 4. Decision Rule Heterogeneity

Corresponding to a model structure, we account for parametric uncertainty by specifying a continuous prior distribution (a multivariate normal) over response parameters (including preference), $\beta^h$. This approach may overcome the limitations of a finite mixture approach that involves a heuristic way of determining the additional number of response segments within a structure (See Kamakura et al. 1996 for details on the heuristic procedure). In addition, our fully Bayesian approach explicitly accounts for heterogeneity in model probabilities across households while Kamakura et al. (1996) assume them to be common across households (i.e., probability of brand-primary decision and that of form-primary decision are common across consumers), though posterior model probabilities vary across consumers.  

4.4 Estimation

The proposed model estimation involves computation of high order multidimensional integrals and therefore makes it practically impossible to use conventional maximum likelihood methods. To circumvent this problem, we use sampling-based methods which involve approximating the joint posterior distribution by obtaining many random draws from the full conditional distributions (rather than directly from the joint posterior distribution) and then basing inference on the empirical distribution of this sample of draws (Details on the full conditional distributions for model parameters are provided in Appendix C). This method is known as substitution sampling.

If all the full conditional distributions are known (i.e., known conjugate distributions such as a normal, a Dirichlet, and a Wishart), then substitution sampling reduces to a procedure known as Gibbs Sampling (Geman and Geman 1984, Gelfand and Smith 1989).

---

8Major criticisms of this type of empirical Bayes approach are overconfidence in prior model probabilities and double counting of data in posterior analysis, i.e., use of the same data to estimate priors and posteriors (Kass and Steffey 1989).
1990). If the full conditional distributions are not completely known (i.e., known only up to a normalizing constant), then a Metropolis-Hastings step can be used (Tierney 1994, Chib and Greenberg 1995). In implementing a sequence of random draws for the proposed model parameters, the \((m+1)\)th step of the substitution sampling involves generating the following draws:

(a) Generate the model indicator \(M\) from a multinomial distribution with a probability vector \(z^h\) for \(h = 1\) to \(H\) where \(z^h\) is the posterior model probability vector (defined in Appendix C) of household \(h\) following each structure.

(b) Generate a \(\lambda^h\) draw from a posterior Dirichlet distribution, \(D(\alpha_1, \ldots, \alpha_s + T^h, \ldots, \alpha_s)\), where \(M = s\) and \(T^h\) is total number of purchases made by household \(h\).

(c) If \(M = s\), generate \(\beta^h_s\) draws from \(p(\beta^h_s | \Sigma^{(m)}, \mu^{(m)}, y^h, M = s)\) for \(h = 1\) to \(H\), using a Metropolis-Hastings algorithm. If \(M \neq s\), generate \(\beta^h_s\) draws from a "pseudoprior", \(p(\beta^h_s | M \neq s)\), defined in Appendix C, i.e., when \(M = s\) we generate from the usual full conditional; when \(M \neq s\) we generate from a "pseudoprior", the linking density (Carlin and Chib 1995).

(d) Generate a \(\mu_s\) draw from \(p(\mu^{(m+1)}_s | \{\beta^{(m)}_s\}, \Sigma^{(m)}_s, \eta_s, C_s)\), for \(s=1,2,3\), using a Gibbs sampler.

(e) Generate a \(\Sigma^{-1}_s\) draw from \(p(\Sigma^{-1(m+1)}_s | \{\beta^{(m)}_s\}, \mu^{(m)}_s, \rho_s, R_s)\), for \(s=1,2,3\), using a Gibbs sampler.

The joint posterior distribution of all unknown parameters is approximated by this sequence of random draws which generates a Markov chain, which becomes stationary after an initial "burn-in" period. Therefore, we discard the initial draws (1st to \(m\)th) from the chain since they only reflect a transient period in which the chain has not
converged to the stationary distribution. A sample of draws obtained after convergence is used to make posterior inferences about model parameters. It should be noted that this sampling-based procedure generates the entire posterior distribution rather than just a point estimate. Given that we expect substantial uncertainty in inferring household parameters with only a small number of observations, it is critical that we should also measure properly the uncertainty of estimates, i.e., a probability interval of the parameter estimates in interest.

In our application, the substitution sampler was run for 15000 iterations and convergence was ensured by monitoring the time-series of the draws. We chose a burn-in length of 10000 iterations (after which the time-series plot stabilized) and retained every third iteration of the remaining 5000 iterations to reduce the serial correlation of the sampled draws, i.e., resort to "thinning the chain" (Geyer 1992, Raftery and Lewis 1995). Therefore, 1000 draws from the posterior distribution of each parameter were used to make inferences.

4.5 Empirical Application

4.5.1 Data

The data for this study are drawn from A.C. Nielsen scanner panel records in the ketchup category for a sample of households in Sioux Falls, South Dakota, for the period 1986-1988. The first 61 weeks were used for initializing model variables, the next 40 weeks of the data (1954 purchases) were used for model calibration, and the last 11 weeks of the data (509 purchases) were used for validation. Households qualified for inclusion in the sample if they made at least one grocery purchase every four weeks over the entire study period and made at least one ketchup purchase both in calibration and validation periods. Out of 592 qualified households, 400 households were randomly selected.
The top seven selling ketchup brands are included in the study. Together, these accounted for 85 percent of category sales in dollars, and 87 percent in units.

4.5.2 Models Estimated

- Pure Sticker Shock Model: Hierarchical Bayes sticker shock model that accounts for parameter heterogeneity only (labeled as HB-ST). This model is estimated setting \( p^h(M = 1) = 1 \) for \( h = 1 \) to \( H \) and skipping steps 1 and 2 described in Section 4.4.

- Pure Reference-Dependent Model: Hierarchical Bayes model of reference-dependent effect that accounts for parameter heterogeneity only (labeled as HB-RD). This model is estimated setting \( p^h(M = 2) = 1 \) for \( h = 1 \) to \( H \) and skipping steps 1 and 2 described in Section 4.4.

- Proposed model: Structural uncertainty-based Hierarchical Bayes model of reference price effects that accounts for both structural uncertainty and parametric uncertainty (labeled as HB-STRD).

4.6 Results and Discussion

In this section we discuss results from the comparison models (HB-ST and HB-RD) and the proposed model (HB-STRD), and compare them in terms of fit statistics such as log-likelihood and hit-rate.

Traditional measures of goodness-of-fit such as the Bayesian Information Criteria (BIC) and Akaike Information Criteria (AIC), which penalize the likelihood for the number of estimated parameters, are not suitable for assessing and comparing Hierarchical Bayes models because these measures are motivated by asymptotic arguments and are, therefore, inappropriate in situations where the number of model parameters varies with the sample size, i.e., when more households are added to the sample,(Carlin and Louis
Chapter 4. Decision Rule Heterogeneity

1996, e.g., Manchanda et. al. 1997). Therefore, we use log-likelihood and hit-rate to compare the models.

Comparison Models: Results of estimating the HB-ST and HB-RD models are presented in Table 4.1. For the calibration period, the HB-RD model outperforms the HB-ST model while for the hold-out period, the results are reversed (except that the hit-rate of the HB-RD model is better). Therefore, it is difficult to favor one model over the other with confidence. If we look at specific parameter estimates of interest, the posterior mean of the sticker shock effect estimate is insignificant (i.e., 0 is included in 95% probability interval) while the loss aversion estimate, measured by the posterior mean difference between the coefficients for both PGAIN and PLOSS, is significant (i.e., 0 is not included in 95% probability interval).\(^9\) In summary, ignoring model uncertainty may lead to the conclusion that either there is no evidence to support the existence of a sticker shock effect (if the HB-ST model is chosen) or that consumers exhibit significant loss aversion (if the HB-RD model is chosen).

Proposed Model: Estimation results for the HB-STRD model is shown in Table 4.2. In an early stage of iterations, simulated model indicator \(M\) hardly took the value of 3, indicating that the mixing rule (within households) is not supported by the data. Therefore, we estimated the proposed model with only two structures, setting \(p^h(M = 3) = 0\) for \(h = 1\) to \(H\).

The HB-STRD model outperforms the comparison models for both calibration and hold-out periods. The HB-STRD model results provide no evidence for the existence of the sticker shock effect while loss aversion\(^10\) still holds. Therefore, it seems that loss aversion phenomenon is robust to heterogeneity correction.

\[^9\]Posterior mean difference between the coefficients for PGAIN and PLOSS is 0.283 (95% probability interval is 0.038 and 0.528). A paired-comparison t-test further confirmed a significant difference.

\[^10\]Posterior mean difference between the coefficients for PGAIN and PLOSS is 0.365 (95% probability interval is 0.160 and 0.570). A paired-comparison t-test further confirmed a significant difference.
Chapter 4. Decision Rule Heterogeneity

The posterior means of household-level model probabilities, $\hat{p}(M = 1 \mid \cdot) = 0.517$ and $\hat{p}(M = 2 \mid \cdot) = 0.483$, give an idea of the proportion of households that follows each decision rule and the range of $p^h(M = 2 \mid \cdot)$ across households (MIN=0.151, MAX=0.835 and standard deviation=0.099) indicates heterogeneity in decision rules across households.

Given insignificant sticker shock effects, we consider households with high $p^h(M = 1 \mid \cdot)$ as having no brand-specific reference price. In contrast, households with high $p^h(M = 2 \mid \cdot)$ can be viewed as having a category-specific reference price and responding more strongly to price losses than to price gains. On average, half the sample households appear to show loss aversion and the remaining households do not seem to have reference prices. In other words, no households seem to have brand-specific reference prices (consistent with the results from Chapter 2) while half of them rely on a category-specific reference price.

4.7 Conclusions

A significant body of research in marketing has investigated the reference price construct and modeled its impact on brand choice decisions via two major competing approaches, the sticker shock and reference-dependent formulations. The two different views on how reference prices affect consumer choice behavior are well-grounded in psychological theory. In addition, Kalyanaram and Winer (1995), in their review paper, interpret the previous research findings as strong support for the existence of reference price effects.

The issue of consumer heterogeneity also has a rich modeling tradition in the marketing literature, and several papers have highlighted the importance of accounting for heterogeneity in obtaining unbiased parameter estimates for state-dependent variables such as reference price used in choice models. Most often, modeling consumer heterogeneity has been limited to accounting for preference and response heterogeneity, i.e.,
allowing for intercepts (preference) and response parameters to vary across households or segments in choice models.

However, less attention has been paid to heterogeneity in decision rules, i.e., different households may have different utility functions (i.e., the sticker shock utility and the reference-dependent utility) as well as different response parameters.

Building on previous research on reference price effects and the notion of model uncertainty, in the current work we develop a Hierarchical Bayes model of reference price effects that account for both model uncertainty and parametric uncertainty, and estimate the model using the MCMC sampling-based methods. As compared to the comparative analysis of reference price models by Briesch et al. (1997), which assumes that all consumers use the same type of reference price, our approach acknowledges that different consumer segments may have different types of reference price, i.e., different decision rules.

Our empirical application of the proposed model to the ketchup category shows that consumers differ not only in their preference and response but also in their decision rules. On average, half the sample households appear to show loss aversion and the remaining households do not seem to respond to brand-specific reference prices. The superiority of the proposed model can be attributed to modeling heterogeneity in decision rules across households and has an added advantage of richer description of consumer choice processes as compared to the comparison models that allow for only one model structure and ignore model uncertainty.

In a broader sense, all kinds of heterogeneity corrections are reflections of the researcher's uncertainty about the true model structure. Therefore, as long as model complexity is manageable and the theoretical underpinnings are compelling, the researcher should seek to build a bigger model. Failure to do this may result in wrong inferences and misunderstanding of the nature of the market.
4.8 Limitations and Future Research

One limitation of the current work is that we operationalize reference price as the price of a brand on the last purchase occasion, following the standard approach in previous research. A more general measure, that includes several past prices, could be used. However, as shown in Chapter 2, such an approach seems to have limited impact on the results, and the last price paid has the strongest results.

In addition, all models do not fit the holdout period as well as the calibration period, indicating a problem of "overfitting". Chiang et al. (1995) also find that the performance of the Hierarchical Bayes model deteriorates in the holdout period. Thus, while Hierarchical Bayes models may provide a good description of the data on which they are calibrated, they may not have a strong forecasting performance. Our findings show that it may be worth making more effort in validating Hierarchical Bayes models in marketing. Moreover, it may be important to develop techniques that can enhance the predictive validity of Hierarchical Bayes model.\textsuperscript{11}

Our empirical analysis showed that the market was described by two distinct segments that follow different decision rules characterized by the sticker shock and reference-dependent models. However, we do not know why each segment uses different decision rules. A tentative explanation about the differences is that a consumer segment may follow the sticker shock model because they have sufficient cognitive capacity and motivation to recall the shelf-prices observed on the last purchase occasion. In contrast, the other segment may follow the reference-dependent model in order to simplify the decision process and use only one, category-specific, reference price. We can investigate this hypothesis further by extending our model, i.e., using covariates to explain the variation of model probabilities ($p^h(M = s)$) across households.

\textsuperscript{11}An exploratory sensitivity analysis shows that as shrinkage toward the prior increases, the calibration fit deteriorates yet the holdout fit enhances and vice versa.
As shown in our empirical results, on average, half the sample households appear to show loss aversion while the remaining households do not respond to (brand-specific) reference prices. Therefore, it seems that loss aversion phenomenon is robust to heterogeneity correction, which is consistent with the results from Murthi and Kalyanaram (1997). In contrast, Briesch et al. (1997) find no loss aversion, in any of the four product categories, after accounting for price response heterogeneity. Bell and Lattin (1996) also show a significant attenuation of loss aversion after heterogeneity correction. As discussed in Chapter 2, their explanation of bias (attenuation of loss aversion) is consistent with our explanation of the underlying cause of a spurious sticker shock effect. Thus, in their work, gains and losses mimic the pattern of price heterogeneity of the underlying segments and play the same role as the sticker shock term does. However, we find that loss aversion persists even after continuously accounting for price response heterogeneity. To understand better the underlying mechanism of bias, if any, in the reference-dependent model, it may be worth deriving analytical results and conducting a simulation study of the kind used in Chapter 2.

Analysis of structural uncertainty is certainly not limited to the reference price research area. It could take many other directions, wherever there are competing hypotheses well grounded in theory and supported by data. For example, one issue on consideration set formation is whether or not one brand can belong to a consideration set without affecting the probability of another brand belonging to the consideration set. "Dependence" among inclusion probabilities means that increased inclusion probability for one brand necessarily decreases the sum of inclusion probabilities for the remaining brands. However, it can be argued that any brand can belong to a consideration set independently. A further generalization may be that one segment of consumers may follow the "dependence" assumption and the other segment may follow the "independence" assumption, depending on their cognitive capacity or willingness to carry brands
in the consideration set. Essentially, this generalization seeks to account for model uncertainty by incorporating competing hypotheses into a single model. We can deal with this by employing the proposed approach. However, it is not always easy to specify all plausible model structures. Guidance is provided by subject-matter considerations such as accepted theories, expert background knowledge, and examinations of conflicting yet complementary empirical results. All this information essentially constitutes model uncertainty. The methods used in this chapter can then be employed to account for this uncertainty and estimate its occurrence.
Table 4.1: Parameter Estimates and Fit Statistics for HB-ST and -RD models

<table>
<thead>
<tr>
<th>Variables</th>
<th>Posterior Means</th>
<th>95% Probability Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>HB-ST</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BLOY</td>
<td>3.602</td>
<td>(3.167, 4.037)</td>
</tr>
<tr>
<td>LBP</td>
<td>0.281</td>
<td>(0.109, 0.452)</td>
</tr>
<tr>
<td>PRICE</td>
<td>−1.642</td>
<td>(−1.818, −1.465)</td>
</tr>
<tr>
<td>RPRICE−PRICE</td>
<td>0.030*</td>
<td>(−0.117, 0.178)</td>
</tr>
<tr>
<td>PROMO</td>
<td>2.429</td>
<td>(2.215, 2.642)</td>
</tr>
<tr>
<td>Calibration LL</td>
<td>−1251.205</td>
<td></td>
</tr>
<tr>
<td>Calibration Hit-rate</td>
<td>76.20%</td>
<td></td>
</tr>
<tr>
<td>Hold-out LL</td>
<td>−571.978</td>
<td></td>
</tr>
<tr>
<td>Hold-out Hit-rate</td>
<td>59.92%</td>
<td></td>
</tr>
<tr>
<td><strong>HB-RD</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BLOY</td>
<td>3.610</td>
<td>(3.112, 4.109)</td>
</tr>
<tr>
<td>LBP</td>
<td>0.108*</td>
<td>(−0.141, 0.357)</td>
</tr>
<tr>
<td>PGAIN</td>
<td>1.438</td>
<td>(1.163, 1.712)</td>
</tr>
<tr>
<td>PLOSS</td>
<td>1.721</td>
<td>(1.520, 1.921)</td>
</tr>
<tr>
<td>PROMO</td>
<td>2.683</td>
<td>(2.268, 3.099)</td>
</tr>
<tr>
<td>Calibration LL</td>
<td>−1212.397</td>
<td></td>
</tr>
<tr>
<td>Calibration Hit-rate</td>
<td>77.07%</td>
<td></td>
</tr>
<tr>
<td>Hold-out LL</td>
<td>−598.751</td>
<td></td>
</tr>
<tr>
<td>Hold-out Hit-rate</td>
<td>60.51%</td>
<td></td>
</tr>
</tbody>
</table>

*: 0 is included in 95% probability interval.
Table 4.2: Parameter Estimates and Fit Statistics for HB-STRD model

<table>
<thead>
<tr>
<th>Variables</th>
<th>Posterior Means</th>
<th>95% Probability Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sticker Shock</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BLOY</td>
<td>3.807</td>
<td>(3.536, 4.077)</td>
</tr>
<tr>
<td>LBP</td>
<td>0.126</td>
<td>(0.001, 0.251)</td>
</tr>
<tr>
<td>PRICE</td>
<td>-1.616</td>
<td>(-1.947, -1.286)</td>
</tr>
<tr>
<td>RPRICE–PRICE</td>
<td>0.160*</td>
<td>(-0.004, 0.324)</td>
</tr>
<tr>
<td>PROMO</td>
<td>2.700</td>
<td>(2.418, 2.982)</td>
</tr>
<tr>
<td>$\hat{p}(M = 1 \mid \cdot)$</td>
<td>0.517</td>
<td>(0.323, 0.711)</td>
</tr>
<tr>
<td><strong>Reference-Dependent</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BLOY</td>
<td>3.730</td>
<td>(3.431, 4.030)</td>
</tr>
<tr>
<td>LBP</td>
<td>-0.021*</td>
<td>(-0.203, 0.161)</td>
</tr>
<tr>
<td>PGAIN</td>
<td>1.589</td>
<td>(1.310, 1.868)</td>
</tr>
<tr>
<td>PLOSS</td>
<td>1.954</td>
<td>(1.673, 2.235)</td>
</tr>
<tr>
<td>PROMO</td>
<td>2.706</td>
<td>(2.484, 2.928)</td>
</tr>
<tr>
<td>$\hat{p}(M = 2 \mid \cdot)$</td>
<td>0.483</td>
<td>(0.289, 0.677)</td>
</tr>
<tr>
<td>Calibration LL</td>
<td>-1171.616</td>
<td></td>
</tr>
<tr>
<td>Calibration Hit-rate</td>
<td>78.25%</td>
<td></td>
</tr>
<tr>
<td>Hold-out LL</td>
<td>-569.840</td>
<td></td>
</tr>
<tr>
<td>Hold-out Hit-rate</td>
<td>60.90%</td>
<td></td>
</tr>
</tbody>
</table>

*: 0 is included in 95% probability interval.
Chapter 5

Conclusions

In this thesis we examined heterogeneity in choice modeling to understand better choice processes and consumer response to the marketing mix.

In the introduction we discussed the kinds of heterogeneity that needs to be accounted for in disaggregate choice analysis, i.e., *parametric uncertainty*, *consideration set heterogeneity*, and *structural uncertainty*. In a broader sense, all kinds of heterogeneity correction are reflections of the researcher's uncertainty about the true model structure. Parametric uncertainty refers to whether or not consumers share common response parameters, and consideration set heterogeneity, whether or not consumers have the same consideration set while structural uncertainty, whether or not consumers have the same utility function (or decision rule). Failure to account for this uncertainty may result in wrong inferences and misunderstanding of choice behavior. In each subsequent chapter, we examined the implications of heterogeneity identified in the introduction and the consequences of ignoring this heterogeneity.

Chapter 2 investigates the impact of unaccounted for price response heterogeneity on estimates of the sticker shock effect. The analysis reveals two conditions under which estimates of the sticker shock effect are biased. First, the estimated model must ignore true differences in price sensitivities among consumers. Second, one, or more, of the other explanatory variables used in the analysis should systematically vary with the underlying price responsiveness. This research highlights the role that heterogeneity in purchase timing can play in this process. Based on this analysis, several predictions about
the nature of parameter bias are made and confirmed via a simulation study. Simulation results also show that, while the existence of extreme price response heterogeneity can, in itself, cause biased estimates of the sticker shock effect, the implied range of price response variation is much larger than observed in practice.

A Hierarchical Bayes version of the nested logit model is proposed and estimated on scanner panel data from two product categories. In contrast to a nested logit model with a sticker shock term, which does not account for price response heterogeneity, the 95% confidence interval of the sticker shock coefficient in the HB-ST model includes the value zero. For both these data sets, therefore, the results do not support the existence of a brand-specific reference price. Household-level category value and price response parameters obtained from the posterior distribution, show that greater responsiveness to category value is associated with higher price sensitivity in the choice decision. Overall, these results provide strong support for our theory of the determinants of sticker-shock bias.

Chapter 3 develops and tests a new "fuzzy" consideration set model that can be estimated with scanner panel. Unlike previous non-fuzzy (well-defined) set approaches which model uncertainty about the sets from which a brand is chosen, the proposed fuzzy set approach directly models the probability of a brand being included in the consideration set. The resulting inclusion probabilities for brands reflect a "fuzzy" consideration set, in the sense that a brand belongs to the consideration set only probabilistically, in contrast to the non-fuzzy consideration sets, in which the sets are well-defined but there is uncertainty about the sets from which a brand will be chosen.

As compared to an existing fuzzy set model (Bronnenberg and Vanhonacker 1996), in which variables have a compensatory influence on consideration, we model consideration as a non-compensatory process. The empirical application to scanner panel data shows that the proposed fuzzy set model outperforms several previous consideration set models.
We apply our modeling approach to examine the impact of advertising on price sensitivity through its influence on consideration set size. In contrast to the experimental findings of Mitra and Lynch (1995), we find no evidence for a positive relationship between consideration set size and price sensitivity. Their hypothesized effects do not appear to be carried over to this particular category. To generalize their findings, further investigation may be necessary with other data set.

Chapter 4 discusses the general problems of ignoring structural uncertainty (decision rule heterogeneity) and develops a Hierarchical Bayes model of reference price effects, in particular, that accounts for heterogeneity in the way that consumers respond to reference prices in making brand choices.

As compared to the comparative analysis of reference price models by Briesch et al. (1997), which assumes that all consumers use the same type of reference price, our approach acknowledges that different consumer segments may have different types of reference price, i.e., brand-specific or category-specific reference price.

Our empirical application of the proposed model to the ketchup category shows that consumers differ not only in their preference and response but also in their decision rules. On average, half the sample households show loss aversion (have a category-specific reference price) while the remaining households do not respond to brand-specific reference prices. In other words, no sample households seem to use brand-specific reference prices (which is consistent with the results from Chapter 2) while half of them rely on category-specific reference price. The proposed model provides a richer description of the consumer choice process than the base models that allow for only one model structure and ignore model uncertainty.

The objective of this thesis was to examine the consequences of ignoring different types of heterogeneity and to develop models that properly account for them to understand better choice processes and consumer response to the marketing mix. Our work points
out some of the limitations of previous studies and shows that it will be necessary to build a model that is sufficiently large to take into account various sources of heterogeneity. While it is generally accepted that scanner panel data analysis should include parametric uncertainty and consideration set heterogeneity, much less attention has been paid to incorporating uncertainty about the model itself. In addition, most often, it is not always easy to pre-specify all plausible model structures. Then we may have to resort to subject-matter considerations such as accepted theories, expert background knowledge, and examinations of conflicting yet complementary empirical results. All this information essentially constitutes model uncertainty.

In conclusion we should seek to build a bigger model as long as model complexity is manageable and theoretical underpinnings are compelling. Ideally, we should incorporate the kinds of heterogeneity mentioned above in a single model so that it may span the range of plausible hypotheses. Otherwise, we may be overconfident of one hypothesis and accept it without questioning its basic assumptions.
Bibliography


---


Appendix A

Derivation of the Conditions Underlying Biased Sticker Shock Effects

Glossary of Symbols

<table>
<thead>
<tr>
<th>Symbols</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_1$</td>
<td>a vector of intercepts in the true model</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>a vector of intercepts in the misspecified model</td>
</tr>
<tr>
<td>$\beta_c$</td>
<td>a common price parameter in the misspecified model</td>
</tr>
<tr>
<td>$\beta_{hi}$</td>
<td>a price parameter for high sensitive segment in the true model</td>
</tr>
<tr>
<td>$\beta_d + \beta_{hi}$</td>
<td>a price parameter for low sensitive segment in the true model</td>
</tr>
<tr>
<td>$\beta_{rp}$</td>
<td>a sticker shock parameter in the misspecified model</td>
</tr>
</tbody>
</table>
Appendix A. Derivation of the Conditions Underlying Biased Sticker Shock Effects

A.1 True model specification

We begin by writing the utility specification of the true model (Equation (2.3)), for a market with n brands, in a general form as,

\[ y = X\beta + u = (x_1:x_2:x_3:x_4) \begin{pmatrix} \alpha_1 \\ \beta_{hi} \\ 0 \\ \beta_d \end{pmatrix} + u. \] (A.1)

\[ X_t = \begin{bmatrix} P_{1t} & (R_{h}^{P} - P_{1t}) & SEG^h \cdot P_{1t} \\ I_t & \vdots & \vdots \\ P_{nt} & (R_{n}^{P} - P_{nt}) & SEG^h \cdot P_{nt} \end{bmatrix} = (x_{1t}:x_{2t}:x_{3t}:x_{4t}), \] (A.2)

With the usual assumptions, \( E(u) = 0, E(uu') = \sigma^2I \) and \( E(X'u) = 0 \). The design matrix for one purchase occasion can be written as:

where, \( I_t \) represents a matrix of the dummy variables of the n brands\(^1\), \( P_{it} \) is price of brand \( i \) on occasion \( t \), \( R_{h}^{P} \) is the reference price of brand \( i \) and \( SEG^h \) is a dummy variable that takes the value 0 if household \( h \) is price-sensitive, and is 1 otherwise. Also, as discussed in the paper, \( \alpha_1 \) is a vector of intercepts, \( \beta_d \) is assumed to be positive and, together with \( \beta_{hi} \), captures the price response of the insensitive segment. The above formulation thus represents the underlying true model consisting of two segments, each with distinctive price sensitivities, but unaffected by reference price since the parameter for the sticker shock term, \( x_3 \), is set to zero.

A.2 Misspecified model

The mis-specified model omits \( x_4 \), i.e., ignores differences in price sensitivity, but includes a sticker shock term, \( x_3 \) and its associated parameter \( \beta_{fp} \). The analyst, therefore,
Appendix A. Derivation of the Conditions Underlying Biased Sticker Shock Effects

estimates the following model:

\[ y = (x_1 : x_2 : x_3) \begin{pmatrix} \beta_1 \\ \beta_c \\ \beta_{rp} \end{pmatrix} + u, \quad (A.3) \]

where \( \beta_1 \) is the vector of parameters corresponding to the brand-specific constants and \( \beta_c \) is the common price coefficient. Let \( X_e \equiv (x_1 : x_2 : x_3) \), denote the matrix of variables used in the estimated model. The estimated parameter vector in an OLS framework is given by

\[
\begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_c \\ \hat{\beta}_{rp} \end{pmatrix} = (X'_e X_e)^{-1} X'_e y \quad (A.4)
\]

Substituting for \( y \) from Equation (A.1) gives

\[
\begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_c \\ \hat{\beta}_{rp} \end{pmatrix} = (X'_e X_e)^{-1} X'_e X \beta + (X'_e X_e)^{-1} X'_e u
\]

\[
= (X'_e X_e)^{-1} X'_e \begin{pmatrix} \alpha_1 \\ \beta_{hi} \\ 0 \\ \beta_d \end{pmatrix} + (X'_e X_e)^{-1} X'_e u
\]

\[
= (X'_e X_e)^{-1} X'_e X_e \begin{pmatrix} \alpha_1 \\ \beta_{hi} \\ 0 \end{pmatrix} + (X'_e X_e)^{-1} X'_e x_4 \beta_d + (X'_e X_e)^{-1} X'_e u \quad (A.5)
\]
The expectations of the estimated parameters can be written as:

\[
E \begin{pmatrix}
\hat{\beta}_1 \\
\hat{\beta}_c \\
\hat{\beta}_{rp}
\end{pmatrix} = \begin{pmatrix}
\alpha_1 \\
\beta_{hi} \\
0
\end{pmatrix} + (X'_e X_e)^{-1} X'_e x_4 \beta_d \\
\text{(A.6)}
\]

The above equation reveals two sources for the bias in \( \hat{\beta}_{rp} \). The first is unaccounted for price response heterogeneity, i.e., a positive value for \( \beta_d \). The second is the term \((X'_e X_e)^{-1} X'_e x_4\), which can be interpreted as the parameter estimates obtained by regressing \( x_4 \) on \( X_e \). If we stack the observations, so that those from the low price-sensitive segment follow those from the high price-sensitive segment, the regression equation can be written as:

\[
x_4 = \begin{pmatrix}
0 \\
0 \\
\vdots \\
P_1 \\
P_2 \\
\vdots
\end{pmatrix} = x_1 \theta_1 + x_2 \theta_2 + x_3 \theta_3 + u \\
\text{(A.7)}
\]

where \( P_1 \) and \( P_2 \ldots \) are the prices of the brands on any purchase occasion.

Note that the elements of \( x_4 \) are a string of zeros in the first rows and a string of positive numbers in the last rows. Under what conditions will the term \((X'_e X_e)^{-1} X'_e x_4\) lead to significance? From the above equation it is clear that this can happen if any of the explanatory variables \( x_1, x_2, \text{ or } x_3 \) mimic the pattern of the dependent variable \( x_4 \). The variable \( x_3 \) (the sticker shock term) is of particular interest to us. Since \( x_3 = RP - x_2 \), we can rewrite the regression equation as follows:

\[
x_4 = x_1 \theta_1 + x_2 (\theta_2 - \theta_3) + RP \theta_3 + u \\
\text{(A.8)}
\]
We see that if the first rows of $\mathbf{RP}$ are lower than the last rows in $\mathbf{RP}$ then this can cause $\theta_3$ to be positive and significant, and thus result in a biased estimate of $\beta_{rp}$. As we argue in the paper, heterogeneity in purchase timing behavior and its relationship to the underlying price sensitivity introduces precisely this pattern in $\mathbf{RP}$, i.e., collectively lower reference prices in the price-sensitive segment than in the price-insensitive segment.

In summary, our derivation establishes two conditions for a spurious sticker shock coefficient. First, there should be unaccounted for price response heterogeneity. In addition, the vector of explanatory variables $\mathbf{X}_e$ should not be orthogonal to $\mathbf{x}_4$, i.e., $\mathbf{X}_e'\mathbf{x}_4 \neq \mathbf{0}$. The magnitude of bias in the sticker shock coefficient is given by the expression:

$$E\left(\begin{array}{c}
\hat{\beta}_1 \\
\hat{\beta}_c \\
\hat{\beta}_{rp}
\end{array}\right) = \left(\begin{array}{c}
\alpha_1 \\
\beta_{hi} \\
0
\end{array}\right) + \left(\begin{array}{c}
\hat{\theta}_1 \\
\hat{\theta}_2 \\
\hat{\theta}_3
\end{array}\right) \beta_d$$  \hspace{1cm} (A.9)

Although not the focus of the present paper, misspecification can also lead to biased estimates of $\beta_1$ and $\beta_c$.

For a more general specification that includes more than 2 segments, the magnitude of the bias is given by a straightforward extension of Equation (A.6) as follows:

$$E\left(\begin{array}{c}
\hat{\beta}_1 \\
\hat{\beta}_c \\
\hat{\beta}_{rp}
\end{array}\right) = \left(\begin{array}{c}
\alpha_1 \\
\beta_{hi} \\
0
\end{array}\right) + (\mathbf{X}_e'\mathbf{X}_e)^{-1}\mathbf{X}_e'\mathbf{x}_4\beta_{d_1} + (\mathbf{X}_e'\mathbf{X}_e)^{-1}\mathbf{X}_e'\mathbf{x}_5\beta_{d_2} + \ldots$$  \hspace{1cm} (A.10)

where $\mathbf{x}_4, \mathbf{x}_5$, and ... are dummy variables corresponding to each price-sensitive segment.
Appendix B

Priors and the Markov Chain Monte Carlo Simulations

B.1 Priors

We specify a multivariate normal prior over household parameters, \( \theta^h \) as follows:

\[
\theta^h \sim MVN(\mu, \Sigma) \tag{B.1}
\]

\( \mu \) and \( \Sigma \) are to be estimated. Following Gelfand et. al. (1990), we put hyper-priors over population parameters, \( \mu \) and \( \Sigma \) as follows:

\[
\mu \sim MVN(\eta, C) \tag{B.2}
\]

\[
\Sigma^{-1} \sim W((\rho R)^{-1}, \rho), \tag{B.3}
\]

where \( \eta = 0, C^{-1} = 0, R = I \), and \( \rho \) is a degree of freedom for Wishart distribution and fixed at 16 (a number of parameters plus one). All these parameters are pre-specified to represent diffuse (non-informative) but proper distribution.

B.2 Full Conditionals and Simulations

B.2.1 Generating \( \theta^{h(m+1)} \)

Given the simulated \( \mu^{(m)} \) and \( \Sigma^{-1(m)} \), we sample household-specific \( \theta^{h(m+1)} \) from the following full conditional distribution:

\[
p(\theta^{h(m+1)} \mid \mu^{(m)}, \Sigma^{(m)}) \propto L(\theta^{h(m+1)})\phi(\theta^{h(m+1)} \mid \mu^{(m)}, \Sigma^{(m)}) \tag{B.4}
\]
where $L(\theta^h_{m+1})=\prod_{t=1}^{T_h} \left[ P_t^h(inc)^{\delta^h_t} (1 - P_t^h(inc))^{1-\delta^h_t} \right]^{\delta^h_t}$, $\delta^h_t = 1$ if household $h$ bought brand $i$ on a store visit at time $t$, and zero otherwise, $\delta^h_t = \sum_\ell \delta^h_{t\ell}$ and $\phi(\cdot | \cdot)$ is a multivariate normal density.

This posterior distribution is known only up to a normalizing constant since it is a product of a normal prior and logit likelihood. Therefore, we adopt the random walk version of the Metropolis-Hastings algorithm (Tierney 1994, Chib and Greenberg 1995, Green 1995). In this version, given the previous draw $\theta^h_{m}$ in the chain, a candidate draw $\theta^h_{c}$ is simulated as:

$$\theta^h_{c} = \theta^h_{m} + u$$

where $u \sim MVN(0, V)$, $V$ is a pre-specified matrix (In our application, following Chiang et al. (1998), we let $V$ be twice the negative inverse of the observed information matrix of the NL-ST likelihood.). Then $\theta^h_{c}$ is accepted as $\theta^h_{m+1}$ in the chain with the following probability:

$$\min \left[ \frac{L(\theta^h_{c}) \phi(\theta^h_{c}) | \mu^{(m)}, \Sigma^{(m)}}{L(\theta^h_{m}) \phi(\theta^h_{m}) | \mu^{(m)}, \Sigma^{(m)}} \right].$$

Therefore, we do not need to know the normalizing constant in simulating draws. If the candidate draw is rejected, we set $\theta^h_{m+1} = \theta^h_{m}$.

### B.2.2 Generating $\mu^{(m+1)}$

Given the simulated $\{\theta^h_{m}\}$ across households and $\Sigma^{-1}(m)$, $\mu^{(m+1)}$ is drawn from the following full conditional distribution (a posterior multivariate normal).

$$p(\mu^{(m+1)} | \{\theta^h_{m}\}, \Sigma^{(m)}, \eta, C) = MVN \left[ D (H \Sigma^{-1}(m) \overline{\theta} + C^{-1} \eta), D \right] \quad (B.5)$$

where $D=(H \Sigma^{-1}(m)+C^{-1})^{-1}$, $\overline{\theta} = \frac{1}{H} \sum_{h=1}^{H} \theta^h_{m}$ and $H$ is the total number of households.
Appendix B. Priors and the Markov Chain Monte Carlo Simulations

B.2.3 Generating $\Sigma^{-1(m+1)}$

Given the simulated $\{\theta^{h(m)}\}$ across households and $\mu^{(m)}$, we simulate $\Sigma^{-1(m+1)}$ from the following full conditional distribution (a posterior Wishart).

$$
\begin{align*}
p(\Sigma^{-1(m+1)} \mid \{\theta^{h(m)}\}, \mu^{(m)}, \rho, R) &= W \left[ \sum_{h=1}^{H} (\theta^{h(m)} - \mu^{(m)}) (\theta^{h(m)} - \mu^{(m)})' + \rho R \right]^{-1}, H + \rho \right].
\end{align*}
$$

(B.6)
Appendix C

Priors and the Markov Chain Monte Carlo Simulations

C.1 Priors and Pseudopriors

C.1.1 Priors on Model Probabilities

\[ \lambda^h \sim D(\alpha_1, \alpha_2, \alpha_3) \]  \hspace{1cm} (C.1)

We use fairly diffuse priors (i.e., \textit{a priori}, we do not favor any model) by setting \( \alpha_s = \frac{1}{\tilde{h}} \). This is equivalent to a prior sample size of 1.

C.1.2 Priors on \( \beta_s^h | M = s \)

We specify independent normal priors on \( \beta_s^h \), for \( s = 1, 2, 3 \) as follows:

\[ \beta_s^h | M = s \sim MVN(\mu_s, \Sigma_s) \]  \hspace{1cm} (C.2)

\( \mu_s \) and \( \Sigma_s \) are to be estimated. Following Gelfand et al. (1990), we put hyper-priors over population parameters, \( \mu_s \) and \( \Sigma_s \) as follows:

\[ \mu_s \sim MVN(\eta_s, C_s) \]  \hspace{1cm} (C.3)

\[ \Sigma_s^{-1} \sim W((\rho_s R_s)^{-1}, \rho_s) \]  \hspace{1cm} (C.4)

where \( \eta_s = 0, C_s^{-1} = 0, R_s = I \), and \( \rho_s \) is a degree of freedom for Wishart distribution and equal to the dimension of parameter vector plus one.\(^1\) All these parameter values are pre-specified to represent diffuse (non-informative) but proper distribution.

\(^1\)The dimension of \( \mu_s \) is twice the dimension of \( \mu_1 \) and \( \mu_2 \) (since variables and intercepts used in both Structure 1 and 2 are required to represent Structure 3), plus one for the mixing parameter \( \gamma^h \).
C.1.3 Pseudopriors on $\beta_s^h \mid M \neq s$

If $M \neq s$ (i.e., if the simulated model indicator points to a structure other than Structure $s$), then the likelihood for $\beta_s^h$ is undefined in our application and posterior updating cannot be implemented. To circumvent this problem and to maintain the Markov chain, we need to specify the "pseudopriors" or linking density from which we can generate $\beta_s^h$ (See Carlin and Chib 1995 for implementational details).

Following Carlin and Chib (1995), as the linking densities, we use estimates of the model-specific posterior distributions $p(\beta_s^h \mid M = s, y^h)$ that can be obtained from individual MCMC runs in which the models with one structure (without structural uncertainty component) are estimated. Matching the linking densities as nearly as possible to the true model-specific posteriors can generate a reasonably well mixing final algorithm since failure to specify reasonable densities will result in slow convergence or making jumps between $M = s$ and $M \neq s$ extremely unlikely in subsequent iterations. It should be noted that we are not using the data to help to select the prior, but only the pseudoprior.

C.2 Full Conditionals and Simulations

C.2.1 Generating Model Indicators

Draw $M$ from a multinomial distribution with $z^h$, the posterior probability vector of household $h$. $(m + 1)$th $z^h$ is updated in the following fashion, from $\lambda^{h(m)}$:

$$z_s^{h(m+1)} = p(M = s \mid \beta^h, y^h) = \frac{\lambda_s^{h(m)} f(y^h \mid \beta_s^h, M = s) \left[ \prod_{i=1}^{3} p(\beta_i^h \mid M = s) \right]}{\sum_{k=1}^{3} \lambda_k^{h(m)} f(y^h \mid \beta_k^h, M = k) \left[ \prod_{i=1}^{3} p(\beta_i^h \mid M = k) \right]} \tag{C.5}$$

where $f(y^h \mid \beta_s^h, M = s)$ is the likelihood for household $h$. 
Appendix C. Priors and the Markov Chain Monte Carlo Simulations

C.2.2 Generating $\lambda^{h(m+1)}$

Generate a $\lambda^h$ draw from the posterior Dirichlet distribution, $D(\alpha_1, \ldots, \alpha_s+T^h, \ldots, \alpha_s)$, where $M = s$ and $T^h$ is total number of purchases made by household $h$.

C.2.3 Generating $\beta^h_s(m+1)$

If $M = s$, given the simulated $\mu_s^{(m)}$ and $\Sigma_s^{-1(m)}$, we sample household-specific $\beta^h_s(m+1)$ from the following full conditional distribution:

$$p(\beta^h_s(m+1) | \mu_s^{(m)}, \Sigma_s^{(m)}, y^h) \propto f(y^h | \beta^h_s(m+1), M = s) \phi(\beta^h_s(m+1) | \mu_s^{(m)}, \Sigma_s^{(m)})$$

(C.6)

where $\phi(\cdot | \cdot)$ is a multivariate normal density.

This posterior distribution is known only up to a normalizing constant since it is a product of a normal prior and logit likelihood. Therefore, we adopt the random walk version of the Metropolis-Hastings algorithm (Tierney 1994, Chib and Greenberg 1995, Green 1995). In this version, given the previous draw $\beta^h_s(m)$ in the chain, a candidate draw $\beta^h_s(c)$ is simulated as:

$$\beta^h_s(c) = \beta^h_s(m) + u_s,$$

where $u_s \sim MVN(0, V_s)$, $V_s$ is a pre-specified matrix. Then $\beta^h_s(c)$ is accepted as $\beta^h_s(m+1)$ in the chain with the following probability:

$$\min \left[ \frac{f(y^h | \beta^h_s(c), M = s) \phi(\beta^h_s(c) | \mu_s^{(m)}, \Sigma_s^{(m)})}{f(y^h | \beta^h_s(m), M = s) \phi(\beta^h_s(m) | \mu_s^{(m)}, \Sigma_s^{(m)})}, 1 \right]$$

Therefore, we do not need to know the normalizing constant in simulating draws. If the candidate draw is rejected, we set $\beta^h_s(m+1) = \beta^h_s(m)$.

\(^2\)In our application, following Chiang et al. (1998), we let $V_s$ be twice the negative inverse of the observed information matrix estimated on the single-segment multinomial logit models of sticker shock ($M = 1$), reference-dependent formulations ($M = 2$), and mixing rules ($M = 3$), respectively, using the maximum likelihood estimation.
Appendix C. Priors and the Markov Chain Monte Carlo Simulations

If \( M \neq s \), we generate \( \beta_s^h \) draws from the pseudoprior, \( p(\beta_s^h \mid M \neq s) \) discussed in C.1.3. Therefore, when \( M = s \) we generate from the usual full conditional; when \( M \neq s \) we generate from a "pseudoprior", the linking density (Carlin and Chib 1995).

C.2.4 Generating \( \mu_s^{(m+1)} \)

Given the simulated \( \{\beta_s^{h(m)}\} \) across households and \( \Sigma_s^{-1(m)} \), \( \mu_s^{(m+1)} \) is drawn from the following full conditional distribution (a posterior multivariate normal).

\[
p(\mu_s^{(m+1)} \mid \{\beta_s^{h(m)}\}, \Sigma_s^{(m)}, \eta_s, C_s) = MVN \left[ D (H \Sigma_s^{-1(m)} \overline{\beta}_s + C_s^{-1} \eta_s), D \right] \quad (C.7)
\]

where \( D = (H \Sigma_s^{-1(m)} + C_s^{-1})^{-1} \), \( \overline{\beta}_s = \frac{1}{H} \sum_{h=1}^{H} \beta_s^{h(m)} \) and \( H \) is the total number of households.

C.2.5 Generating \( \Sigma_s^{-1(m+1)} \)

Given the simulated \( \{\beta_s^{h(m)}\} \) across households and \( \mu_s^{(m)} \), we simulate \( \Sigma_s^{-1(m+1)} \) from the following full conditional distribution (a posterior Wishart).

\[
p(\Sigma_s^{-1(m+1)} \mid \{\beta_s^{h(m)}\}, \mu_s^{(m)}, \rho_s, R_s) = W \left[ \sum_{h=1}^{H} (\beta_s^{h(m)} - \mu_s^{(m)})(\beta_s^{h(m)} - \mu_s^{(m)})^t + \rho_s R_s \right]^{-1}, H + \rho_s \]
\]
\[ (C.8) \]