The Pennsylvania State University

The Graduate School

The Mary Jean and Frank P. Smeal College of Business Administration

A HIERARCHICAL BAYESIAN FINITE MIXTURE MULTIDIMENSIONAL SCALING APPROACH FOR ACCOMMODATING STRUCTURAL AND PREFERENCE HETEROGENEITY IN THREE WAY PREFERENCE DATA

A Thesis in

Business Administration

by

Joonwook Park

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

May 2007

The thesis of Joonwook Park was reviewed and approved* by the following:

Wayne S. DeSarbo
Mary Jean and Frank P. Smeal Distinguished Professor of Marketing
Thesis Advisor
Chair of Committee

Arvind Rangaswamy
Anchel Professor of Marketing

Duncan K. Fong
Professor of Marketing and Statistics

John C. Liechty
Associate Professor of Marketing and Statistics

Joseph L. Schafer
Associate Professor of Statistics

Hans Baumgartner
Professor of Marketing and Charles & Lillian Binder Faculty Fellow
Head of the Department of Marketing

*Signatures are on file in the Graduate School

ABSTRACT

Latent Class Multidimensional Scaling Models (LCMDS hereafter) have been widely used in the marketing and psychometric literature to capture consumers' preference heterogeneity and market structure simultaneously. In LCMDS analysis, two types of utility models, the vector and the ideal point model, have been used to represent consumers' preferences. These models, however, lead to different representations of market structure and consumers' heterogeneity. The ideal point model, a more general case of the vector model, often suffers from degenerate solutions. Recent research suggests that such ideal point degenerate solutions can be results of mixing sample of consumers that utilize different underlying utility functions. As such, we propose a Hierarchical Bayesian Finite Mixture MDS approach to take both preference and structural heterogeneity into account. We show that both the ideal point only model and the vector only model are special cases of the proposed model. We then discuss the Markov Chain Monte Carlo sampling utilized to generate the posterior distributions of the unknown model parameters. We then apply the model to simulated data, and an application from the pharmaceutical industry in the area of antidepressant prescriptions by medical doctors over time. We find that the model with structural and preference heterogeneity outperforms models without structural heterogeneity. Results suggest that ideal point segments and vector segments can be distinguished by their average prescription volume and preference variance. Vector segments exhibit higher prescription volume despite the fact that they consist of a small number of physicians than ideal point

segments. Furthermore, vector segments show higher preference variance across brands. Together, this implies that the proposed approach can provide an efficient way of targeting physicians given that high volume physicians draw more attention for detailing effort but are less responsive to this marketing effort than low volume physicians.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

The completion of this dissertation was made possible through the support and encouragement of many individuals, and I would like to take this opportunity to acknowledge them.

My first, and most earnest, acknowledgment must go to my advisor and chair of my Committee, Dr. Wayne S. DeSarbo for his advice, guidance, direction, and endless enthusiasm for the subject of this dissertation. Dr. DeSarbo has been instrumental in ensuring my academic, professional, financial, and moral well-being ever since I came to Pennsylvania State University. In every sense, none of this work would have been possible without him.

I must also recognize the influence of the other members of my dissertation committee, Dr. Arvind Rangaswamy, Dr. Duncan K. H. Fong, Dr. John C. Liechty, and Dr. Joseph Schafer. Each brought his own insight to the process, and changed the way I see the research process. I would like to thank Dr. Rangaswamy for his generous financial support and his input for the managerial aspect of this dissertation. Dr. Fong and Dr. Schafer put forth much effort concerning the improvement of the dissertation. My special thanks should go to Dr. Liechty for his insightful input to this dissertation. He has inspired me with respect to Bayesian statistics and has spent much time to discuss the technical development of the paper.

Far too many people to mention individually have assisted in so many ways during my work at Pennsylvania State University. They all have my sincere gratitude. In particular, I would like to thank Crystal Scott, Selin Atalay, Girish Mallapragada, and Rui Wang among many other Ph.D. students. They have shared their insights, experiences, and joys together. I am largely indebted to their friendship.

I must also acknowledge the help of professors in the Marketing Department. Dr. Margaret Meloy and Dr. Bill Ross always opened their doors and mentored me regarding different matters. I would like to I thank the Marketing Department staff, especially Ms. Stephanie Ironsides. She made coming to work bearable on bad days and exhilarating on good days.

A penultimate thank-you goes to my wonderful parents. For always being there when I needed them most, and always supporting my decision of pursuing graduate study, they deserve far more credit than I can ever give them.

Finally, I must acknowledge my family. My children, Kevin, and Jina, have endured long excuses from their father for not spending time together. They kept me grounded during this long process and gave me a place to escape to when the work became overwhelming.

My most heartfelt acknowledgment must go to my wife, Mina, for her companionship during every step of this process. I honor her for her love without which this would not have been possible. Her support and encouragement have turned my journey through graduate school into a pleasure. For all that, and for being everything I am not, she has my everlasting love.

DEDICATION

To

Kevin, and Jina Park

Mina Jung Park

Dooyong Park, Soonhee Kim, Dongchul Jung, and Kaenam Kim

Chapter 1

INTRODUCTION

Multidimensional Scaling (MDS) refers broadly to a plethora of spatial models utilized to obtain multidimensional spatial representations for various types of data, including proximity or dominance[1] data. MDS procedures have been one of the main vehicles for product positioning in marketing (see Urban and Hauser 1993; Wind 1982). The last two decades have seen major developments in the area of Latent Class Multidimensional Scaling methods (hereafter referenced as LCMDS) that simultaneously derive market structure, positioning, and segmentation (Andrews and Manrai 1999; Bőckenholt and Bőckenholt 1991; Bőckenholt and Gaul 1989; Chintagunta 1994; DeSarbo, Howard, and Jedidi 1991; DeSarbo and Jedidi 1995; DeSarbo, Manrai, and Manrai 1994; DeSarbo, Ramaswamy, and Lenk 1993; Sinha and DeSarbo 1998; Wedel and DeSarbo 1996; Winsberg and Soete 1993). The popularity of many LCMDS analyses stems from their ability to represent jointly the underlying brand positions and the preferences that unknown consumer segments exhibit from various types of data. Such brand positions can provide valuable market competition information so that brand managers are able to implement successful marketing strategies.

In LCMDS analysis of dominance data, two distinct types of utility models have traditionally been used to represent consumers' preferences: Slater's (1960) and Tucker's

---

[1] Dominance data represents any quantifications defined over sets of stimuli that quantify relative superiority of one stimulus over others. An example is preference rating data.

(1960) vector model and Coombs' (1964) ideal point model. Although both models assume that consumers reach their preferences by considering a multidimensional set of brand characteristics, these models have different consumer preference assumptions (DeSarbo, Young, and Rangaswamy 1997). The goal of the ideal point model, also known as the distance-based unfolding model, is to recover the preferences of consumers by representing both consumers and brands as coordinates such that the distance between brands and consumers decreases as consumers' preferences increase[2]. Consumers' coordinates, known as ideal points, specify the ideal combination of brand characteristics for these consumers (Wedel and Kamakura 2000). The underlying assumption of the simple ideal point or unfolding utility model is that preference decreases as a brand coordinate is positioned farther away from an ideal point in any direction. In the simple ideal point utility in two dimensions, the iso-preference contours are circles of equal radii around the ideal point. This ideal point model is also perceived as a more generalized case of the vector model (Carroll 1972) but often suffers from degenerate solutions that hinder marketers from interpreting the derived solutions.

The vector model represents consumers by vectors and brands by coordinates. Here, brand coordinates represent the relative positions of brands in space, and consumers' preferences are recovered by the orthogonal projection of the brands' coordinates onto vectors. As such, higher values of projection indicate higher preferences, and the length of vector shows the relative magnitude of preference. The

---

[2] This distance often implies Euclidean distance (or squared Euclidean distance) on space, and iso-preference contours for the simple ideal point model are families of concentric circles (or spheres, or hyper-spheres for more than two dimensions)

underlying assumption of the vector utility model is “the more the better”: brands positioned farther out in the direction of the segments’ (or individuals’) vector are more preferred. Therefore, the iso-preference contours in two dimensions are straight lines perpendicular to the vector. Figures 1.1 and 1.2 illustrate an example of the simple ideal point and the vector LCMDS model in two dimensions respectively.

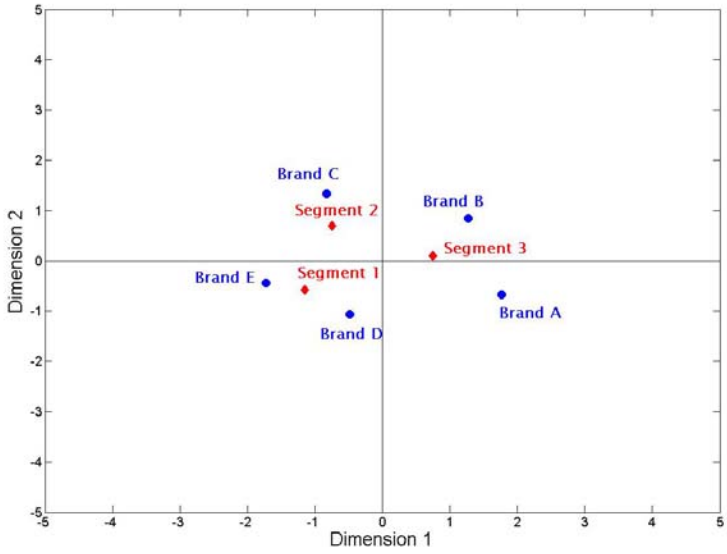

Figure 1.1: An Ideal Point LCMDS Model

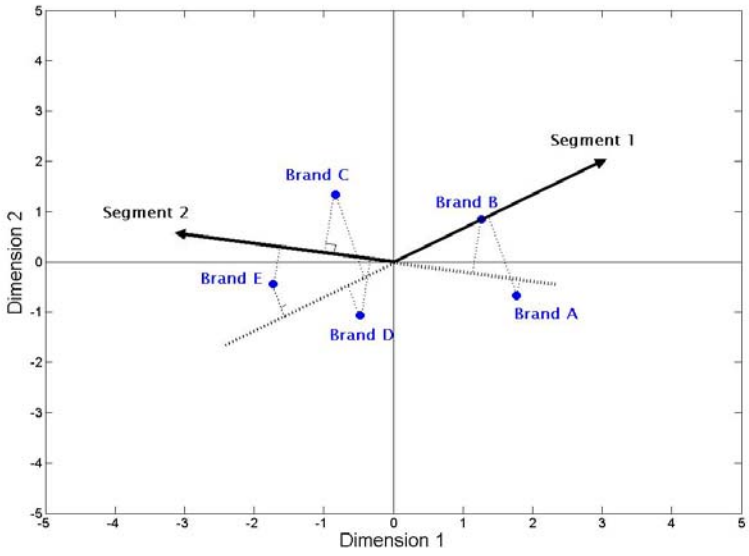

Figure 1.2: A Vector LCMDS Model

Although such spatial models have been separately applied in various marketing applications, little attention has been given to comparative structural heterogeneity in contrast to the vast literature on preference heterogeneity in LCMDS analysis. Here, preference heterogeneity refers to the difference in preference across consumers (Yang and Allenby 2000). Preference heterogeneity is normally described either by exogenous covariates, such as demographics (e.g., household income), or by the independent draws from a mixing distribution (i.e., random-effects models or finite mixture models). Structural heterogeneity refers to differences in the structure of underlying decision processes (Kamakura, Kim, and Lee 1996). As such, implicit in the assumption of existing LCMDS methods is that all consumers are homogeneous in the manner in which

they form their preference. Recent research, however, suggests that this may not be the case. Deun, Groenen, Heiser, Busing, and Delbeke (2005) contend that there is a close connection between the vector model and degenerate solutions frequently encountered in the ideal point model, and that those ideal points farther from the centroid of brands' coordinates can be replaced by vectors without altering consumers' preference order (Deun, Groenen, Heiser, Busing, and Delbeke 2005). This implies that the vector model and the ideal point model need to be considered as structurally heterogeneous, and that a generalized approach which can explicitly accommodate "mixtures of vector and unfolding (ideal point) representations" (DeSarbo and Carroll 1985) would be an important contribution to the LCMDS literature.

Therefore, we propose a generalized LCMDS model that explicitly takes both structural and preference heterogeneity into account. Specifically, we propose a new Bayesian LCMDS model which combines both the vector model and the ideal point model in a generalized framework for modeling preference or other forms of dominance data. The proposed model has several merits. First, this model explicitly allows for mixtures of the vector and the ideal point model, thereby accounting for structural heterogeneity. As such, both the vector and the ideal point model can be considered as special cases of the proposed model. Secondly, preference heterogeneity is accommodated via a finite mixture based method which enables one to identify unknown market segments that exhibit different perception toward brands. Lastly, this model employs a Hierarchical Bayesian finite mixture method that incorporates uncertainty related to the model parameter estimates.

Chapter 2

THEORETICAL DEVELOPMENT

Carroll (1972) has argued that the vector model is a special case of the ideal point model. That is, if one moves an ideal point farther and farther in an optimal direction from the origin, while brands' coordinates remain fixed, then the rank order of distances from the ideal point would be asymptotically identical to that of projections of brands' coordinates along the vector (Carroll 1980). This implies that a brand located farther from the origin would be more preferred; hence, "more is better." Furthermore, this suggests that the interpretations of these models should be similar if one allows infinite ideal points. Here, the recovery of consumers' preferences should be identical from these models. Unfortunately, this is not the case when ideal points are finite. Figure 2.1 illustrates a hypothetical case where interpretations of consumers' preference are not consistent across these models. For easy of exposition, we assume that there are four brands denoted by A, B, C, and D, and one ideal point segment. Note that each circle represents an iso-utility contour. Therefore, the recovered preference order from the ideal point representation is A > B > C > D. Now, assume that we draw a vector from origin to the coordinate of the ideal point and that we draw a perpendicular line from coordinates of brands onto this vector. The recovered preference order from this vector representation is A > C > B > D. This hypothetical case illustrates that the interpretation from the vector model is not matching with that from the ideal point model.

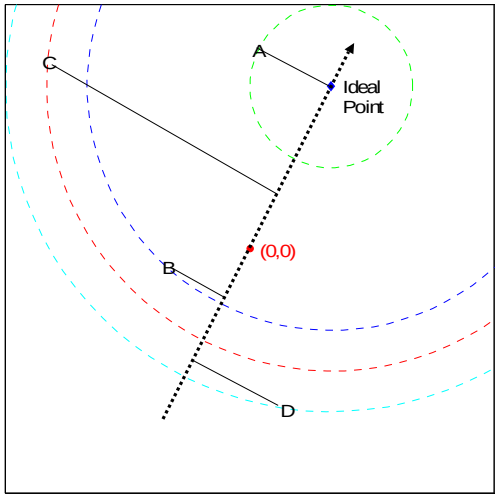

Figure 2.1: Vector and Ideal Point Model Representation of the Recovery of Preference

The ideal point model often suffers from degenerate solutions which hinder the interpretation of the result from the ideal point model. Heiser (1989) defines the degenerate solutions as "extremely uninformative solutions with good or even perfect fit." For instance, one of the common degenerate solutions occurs when the two sets of coordinates (i.e., ideal points and brands) lie in concentric circles where the ideal points encompass the brand coordinates. This pattern can be characterized as "the objects[3] are at an (almost) equal distance to the subjects" (Deun, Groenen, Heiser, Busing, and Delbeke 2005) and is called object point (or object implosion) configuration (Heiser 1989). Degenerate solutions can be classified into four different patterns: (1) Object point

[3] We interchange objects with brands and subjects with consumers.

configuration where brands are densely located at origin (or any point), and ideal points encompass brand coordinates; (2) object circle configuration where ideal points are densely located at a certain point (e.g., origin), and brands coordinates encompass ideal points; (3) one-plus-one configuration where brand coordinates collapse into one location, and ideal points also collapse into another location; and (4) partial degenerate configuration where some ideal points are located close to brands' coordinates while other ideal points are located at points that are far away from any coordinates of brands. Figure 2.2 illustrates various types of degenerate solutions.

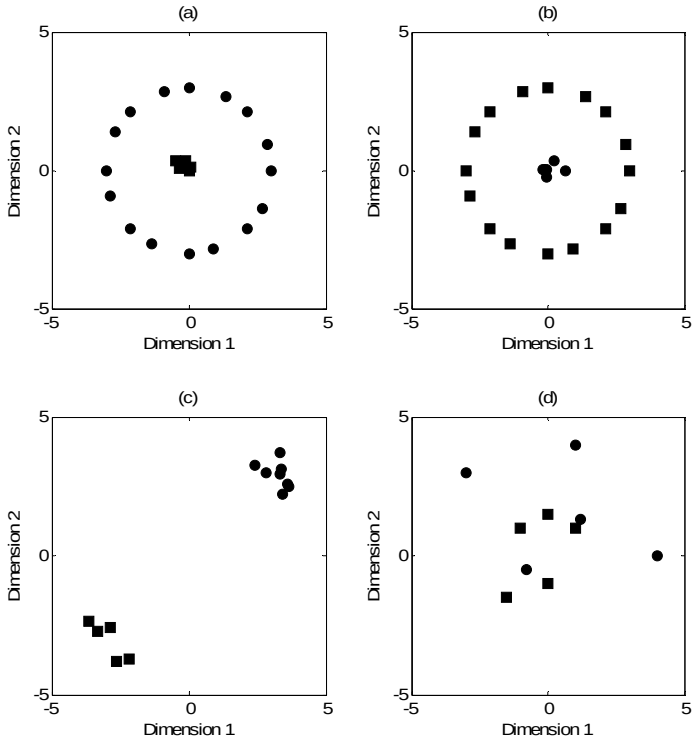

Figure 2.2: Types of Degenerate Solutions: (a) Object point configuration, (b) Object circle configuration, (c) One-plus-one configuration, and (d) Partial degenerate configuration

Note: squares represent brand coordinates, and circles represent ideal points

The reason for these degenerate solutions is not clearly understood. For instance, DeSarbo and Rao (1984, 1986) argue that a particular form of error structure or flat response surface might cause degenerate solutions. In algorithmic approaches that tend to minimize a badness of fit function, it is argued that transformations of the data (either ordinal or interval) may equalize them to a single intercept term, which can produce one-plus-one degenerate configuration (Deun, Groenen, Heiser, Busing, and Delbeke 2005). Furthermore, it is contended that there is a close connection between degenerate solutions and the vector model. De Leeuw (1983) showed that unfolding algorithms minimizing Kruskal's Streess-2 tend to generate degenerated solutions that optimize the fit of the vector model. Based on this study, Deun, Groenen, Heiser, Busing, and Delbeke (2005) propose a procedure for the ideal point model with degenerate solutions such that ideal points farther from the centroid of brands' coordinates are replaced by vectors. Specifically, they propose a procedure in the following fashion: (1) a researcher investigates a derived solution from the ideal point model; (2) if an object-circle type degenerate solution is observed, then he/she sets a minimum distance, $d_{\min}$; (3) he/she replaces ideal points located outside of $d_{\min}$ with vectors; and (4) the researcher repeats procedures (1) – (3) until the recovered preference based on a combination of ideal points (the points falling in or on the circle with radius $d_{\min}$) and vectors (the remaining points falling outside the circle with radius $d_{\min}$) has at least the same recovered preference orders as in the original ideal point solution. Deun et al. (2005) show that they can recover the original preferences of consumers by replacing some ideal points in degenerate solutions with vectors.

Application of this strategy is, however, somehow questionable since (1) this procedure employs an arbitrary search strategy; (2) an employment of this strategy is at best post-hoc; (3) application to a large number of consumers may be hindered due to the algorithmic approach; (4) this algorithmic approach does not provide a formal test for the number of dimensions and significance test for estimated parameters (Wedel and Kamakura 2000); and (5) this procedure requires that the recovered preference order by the replaced vectors should be the same as or better than that of the original ideal point model estimated, which implicitly assumes that the utility derived from the ideal point model is always better than that from the vector model.

Several other remedies have been developed to help researchers cope with these degenerate solutions (DeSarbo, Young, and Rangaswamy 1997). The simplest approach is to use an external unfolding analysis in which the coordinates of brands are obtained a priori and remain fixed during the estimation procedure (Carroll 1972; Kamakura 1991). Other approaches are the imposition of restrictions on the configuration of the derived joint space (Heiser 1981) or the use of a weighted loss function (DeSarbo and Carroll 1985; DeSarbo and Rao 1986; DeSarbo and Rao 1984). DeSarbo, Young, and Rangaswamy (1997) devised indices that measure the degree of degeneracy for an ideal point model and developed a procedure for non-metric data (e.g., rank-ordered data) in which each consumer's ideal point is restricted at a weighted average of the coordinates of brands. For a review of alternative remedies for degenerate solutions, see Kim, Rangaswamy, and DeSarbo (1999).

We contend that an explicit approach that acknowledges both preference and structural heterogeneity can explain consumers' heterogeneity from a more theoretical

perspective. Researchers have contended that the evaluation process is more contingent upon consumers' knowledge structures and/or decision strategies (Bettman, Luce, and Payne 1998; Sujan 1985). For instance, more knowledgeable consumers are more likely to evaluate product information via an attribute-based process, especially when incoming information is discrepant to their prior category knowledge (Sujan 1985). On the contrary, novice consumers would be more likely to rely on more simplistic criteria in making judgments. Furthermore, some consumers may process a different amount of information (e.g., selective processing) while others may process an equal amount of information for each attribute or alternative (e.g., consistent processing) (Bettman, Luce, and Payne 1998). For example, when a consumer makes judgment of automobiles, a particular attribute (e.g., safety) may be the most salient attribute to him/her, and this consumer may make his/her decision solely based on the safety features, while other consumers may consider other attributes such as horsepower, reliability, and price with equal weight.

In the marketing literature, several researchers have paid attention to structural heterogeneity underlying choice processes (Gensch 1987; Gilbride and Allenby 2004; Jedidi and Kohli 2005; Kamakura, Kim, and Lee 1996; Yang and Allenby 2000). Gensch (1987) showed the existence of consumers using different choice rules. Assuming that a model with structural heterogeneity would show better prediction for consumers' choices, he compares the predictions of the logit model, the maximum likelihood hierarchical model (MLH), and a hybrid model where he applied the logit model to more knowledgeable consumers and the MLH to less knowledgeable consumers. He found that

the hybrid model that reflects structural heterogeneity outperforms models without structural heterogeneity in the prediction of actual choice.

Kamakura, Kim, and Lee (1996) proposed a model with a two-stage hierarchical choice process in a nested logit framework. In their model, there were two types of hierarchical choice processes tested: Brand type segments were assumed to choose a brand first, and then a product-form in a hierarchical fashion, whereas Product Form type segments first choose the form of product and then the brand subsequently. By allowing preference heterogeneity within segments, Kamakura and colleagues (1996) jointly modeled these two types of heterogeneity in a finite mixture model framework. Their results show the model with both structural and preference heterogeneity dominates models that do not consider these sources of heterogeneity simultaneously in both model fit and validation.

Recently, Gilbride and Allenby (2004) compare three distinctive choice screening rules: compensatory rule, conjunctive rule, and disjunctive rule. The compensatory rule is defined as one where the deterministic part of the utility of an alternative should exceed a threshold value to be acceptable. The conjunctive rule and the disjunctive rule are more related to attribute-based processing. The conjunctive rule refers to one where the utility of each attribute should exceed the threshold at its own attribute level, whereas the disjunctive rule is defined as one where at least one of the attribute-level utilities is acceptable. They show that a model with the conjunctive rule dominates other choice process rules empirically, which is closely matched by a structural heterogeneity model where the conjunctive and the disjunctive rule are combined. Jedidi and Kohli (2005) propose a two stage model in which a subset-conjunctive rule that is a generalized

decision rule of the disjunctive and the conjunctive rule is applied in the first (consideration) stage, and a finite mixture model is applied to account for consumers' heterogeneity in the second (choice) stage. Here, the subset-conjunctive rule is defined as one where an alternative is acceptable to consumers if at least a subset of attributes is acceptable (e.g., s out of m attributes where $1 \leq s \leq m$). As such, both the conjunctive ($s = m$) and the disjunctive rule ($s = 1$) are special cases of this subset-conjunctive rule. They show that their two stage model outperforms other models (i.e., Dogit model) in terms of model fit and predictability. Yang and Allenby (2000) also propose a model that incorporates household level heterogeneity, as well as structural heterogeneity, in a hierarchical Bayesian fashion for credit card adoption behavior. They assume that structural heterogeneity can be captured whether consumers compare the new card offer to the best card they are holding, or whether they compare the best feature regardless of the card. Their study shows somewhat mixed results. That is, although a model with both structural and preference heterogeneity had worse fit than a model with only preference heterogeneity in the calibration sample, the former showed a better fit in the hold-out sample.

In summary, previous research suggests that consumers' preferences may be better represented by incorporating both structural and preference heterogeneity in a model. In LCMDS analysis, this means that some consumers' preferences may be better represented by the ideal point model while others' preferences would be better represented by the vector model. In the next chapter, we propose a generalized LCMDS model that incorporates both preference and structural heterogeneity in a Bayesian finite mixture framework.

Chapter 3

THE PROPOSED BAYESIAN FINITE MIXTURE LCMDS MODEL

3.1 Model Development

In this chapter, we first define the vector and ideal point LCMDS models, which we then generalize to develop the proposed model. We then explain how these two models can be incorporated into one general framework to accommodate structural heterogeneity. Next, we briefly describe the Bayesian framework including prior specifications and an estimation procedure using a Markov Chain Monte Carlo method. Detailed descriptions of full conditional distributions are also provided.

For i = 1,…, I consumers who make preference judgments toward j = 1,…, J brands in r = 1,…, R situations (e.g., time, experimental treatments, or consumption occasion), let Δ_{ijr} denote preference rating (or dominance score) for brand j by consumer i in the r-th situation; t = 1,…, T unknown dimensions; $s_1 = 1,....,S_1$ unknown segments obeying the vector model; and $s_2 = 1,....,S_2$ unknown segments obeying the ideal point model. Following the finite mixture model framework, it is assumed that there are *S* ($S = S_1 + S_2$) unknown segments of which S_1 segments obey the vector model, and S_2 segments obey the ideal point model. Next, we define the vector and the ideal point utility as:

$$U_{ijr|s_1} = \sum_{t=1}^{T} x_{jt} v_{s_1 t} + b_{s_1} + e_{ijr|s_1} , \quad (3.1)$$

$$U_{ijr|s_2} = c_{s_2} \sum_{t=1}^{T} \left(x_{jt} - w_{s_2 t} \right)^2 + d_{s_2} + e_{ijr|s_2} , \quad (3.2)$$

where $e_{ijt|s_1} \sim N(0, \delta_{s_1}^2)$ and $e_{ijt|s_2} \sim N(0, \delta_{s_2}^2)$. Here, $U_{ijr|s_1}$ and $U_{ijr|s_2}$ represent consumer i's utility toward brand j in situation r given that the consumer belongs to the vector model segment or the ideal point model segment respectively. Without much loss of generality, we assume that $\delta_{s_1}^2 = \delta_{s_2}^2 = 1$.

In equations 3.1 and 3.2, x_{jt} represents a brand j's coordinate in the t-th dimension, and b_{s_1} and d_{s_2} represent additive constants for the vector model and the ideal point model respectively. $v_{s_1 t}$ represents the t-th coordinate of segment s_1's vector, and $w_{s_2 t}$ represents the t-th coordinate of the ideal point for segment s_2. Finally, c_{s_2} is a scale parameter for the squared Euclidean distance between brand coordinates and ideal points of Segment s_2. As such, equations 3.1 and 3.2 represent typical vector and ideal point utility functions respectively. Note that an extension of the simple ideal point model in equation 3.2 is a weighted ideal point model in which each segment has its own weight for dimension t (DeSarbo and Carroll 1985; DeSarbo and Rao 1984; Wedel and DeSarbo 1996). Similarly, the scale parameter c_{s_2} is introduced to weight the squared Euclidean distance at the segment level. As indicated in GENFOLD2 (DeSarbo and Rao 1986; DeSarbo and Rao 1984), a positive sign of scale parameter c_{s_2} indicates an anti-ideal point

when the data represents preference[4]. Thus, this scale parameter c_{s_2} is typically constrained to be strictly negative to represent the ideal point model. We impose a restriction for this scale parameter in the prior distribution section. It should be noted that if $c_{s_2} < -1$, then the derived Euclidean distance is shrunken and if $-1 < c_{s_2} < 0$, then the derived Euclidean distance is stretched to a certain degree in equation 3.2. This would cause inconsistency between the derived Euclidean distance and the interpretation of derived map based on the parameters x_{jt} and $w_{s_2 t}$.

Given the utility specified in equations 3.1 and 3.2, the individual conditional likelihood functions for the vector and the ideal point models are constructed as follows:

$$L_{i|V} = \sum_{s_1=1}^{S_1} \left(\prod_{r=1}^{R} \prod_{j=1}^{J} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} (\Delta_{ijr} - \sum_{t=1}^{T} x_{jt} v_{s_1 t} - b_{s_1})^2 \right) \right) \gamma_{s_1}, \tag{3.3}$$

$$L_{i|IP} = \sum_{s_2=1}^{S_2} \left(\prod_{r=1}^{R} \prod_{j=1}^{J} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} (\Delta_{ijr} - c_{s_2} \sum_{t=1}^{T} (x_{jt} - w_{s_2 t})^2 - d_{s_2})^2 \right) \right) \gamma_{s_2}, \tag{3.4}$$

where $L_{i|V}$ and $L_{i|IP}$ represent conditional individual level likelihoods given that the observations from an individual i belongs to the vector model or the ideal point model respectively. Here, parameters γ_{s_1} and γ_{s_2} are unknown membership probabilities for the vector and the ideal point model segments such that $\sum_{s_1=1}^{S_1} \gamma_{s_1} = 1$ and $\sum_{s_2=1}^{S_2} \gamma_{s_2} = 1$. Following the standard data augmentation procedure (Diebolt and Robert 1994; Tanner and Wong

[4] In the case of dispreference data, a negative sign of scale parameter c_{s_2} indicates an anti-ideal point.

1987), we introduce latent variables ξ_{is_1} and ξ_{is_2} such that $\Pr(\xi_{is_1} = s_1) = \gamma_{s_1}$ and $\Pr(\xi_{is_2} = s_2) = \gamma_{s_2}$. Therefore equations 3.3 and 3.4 can be rewritten as:

$$L_{i|V} = \sum_{s_1=1}^{S_1} \left(\prod_{r=1}^{R} \prod_{j=1}^{J} f_{v_{s_1}}(\Delta_{ijr} \mid x_{jt}, v_{s_1 t}, b_{s_1}) \right) I(\xi_{is_1} = s_1), \tag{3.5}$$

$$L_{i|IP} = \sum_{s_2=1}^{S_2} \left(\prod_{r=1}^{R} \prod_{j=1}^{J} f_{IP_{s_2}}(\Delta_{ijr} \mid x_{jt}, w_{s_2 t}, c_{s_2}, d_{s_2}) \right) I(\xi_{is_2} = s_2), \tag{3.6}$$

where $I(\,)$ is an indicator function, and

$$f_{v_{s_1}}(\Delta_{ijr} \mid x_{jt,} v_{s_1 t}, b_{s_1}) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} (\Delta_{ijr} - \sum_{t=1}^{T} x_{jt} v_{s_1 t} - b_{s_1})^2 \right)$$

$$f_{IP_{s_2}}(\Delta_{ijr} \mid x_{jt,} w_{s_2 t,} c_{s_2}, d_{s_2}) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} (\Delta_{ijr} - c_{s_2} \sum_{t=1}^{T} (x_{jt} - w_{s_2 t})^2 - d_{s_2})^2 \right).$$

Thus, equations 3.5 and 3.6 are not much different from existing LCMDS models. We now incorporate structural heterogeneity by introducing another latent indicator variable such that $\Pr(\chi_i = 0) = \phi_0$ and $\Pr(\chi_i = 1) = 1 - \phi_0$. This latent variable determines the probability that an individual i belongs to either a vector model segment (i.e., ϕ_0) or an ideal point model segment (i.e., $1 - \phi_0$). The resulting unconditional likelihood is:

$$L = \prod_{i=1}^{I} \left[\begin{array}{l} \left(\prod_{s_1=1}^{S_1} \prod_{r=1}^{R} \prod_{j=1}^{J} f_{v_{S_1}}(\Delta_{ijr} \mid x_{jt}, v_{s_1 t}, b_{s_1}) \right)^{I(\chi_i = 0, \xi_{is_1} = s_1)} \\ \times \left(\prod_{s_2=1}^{S_2} \prod_{r=1}^{R} \prod_{j=1}^{J} f_{IP_{s_2}}(\Delta_{ijr} \mid x_{jt}, w_{s_2 t}, c_{s_2}, d_{s_2}) \right)^{I(\chi_i = 1, \xi_{is_2} = s_2)} \end{array} \right], \tag{3.7}$$

where $I(\chi_i = 0, \xi_{is_1} = s_1)$ represents a joint indicator function (i.e., $I(\chi_i = 0, \xi_{is_1} = s_1) = I(\chi_i = 0) I(\xi_{is_1} = s_1))$. It should be noted that an individual is assigned to one and only one segment due to this joint indicator function, and that this assignment is different from a "fuzzy" classification in which an individual has a density of segment assignment (e.g., Kamakura and Russell 1989). Thus, the individual level likelihood is obtained either from one of the vector model segments or from one of the ideal point model segments in equation 3.7. Note that the specification of equation 3.7 is a generalized form of the vector and the ideal point model. If no individual is assigned to vector model segments, then this model becomes the ideal point model, and vice versa[5].

To summarize, we have postulated a LCMDS model with both preference heterogeneity given by equations 3.5 and 3.6 and structural heterogeneity given by equation 3.7. To estimate the proposed model, we employ a Bayesian finite mixture approach. From the Bayesian perspective, it is straightforward to construct a finite mixture model in a hierarchical Bayesian fashion by specifying appropriate priors. Recent developments in Bayesian statistics enable researchers to use simulation-based methods known as Markov Chain Monte Carlo methods (hereafter referenced as MCMC) to generate random deviates from the posterior distribution without requiring analytic integration (Chib 2002; Gelfand and Smith 1990; Gilks, Richardson, and Spiegelhalter 1996; Hastings 1970; Metropolis, Rosenbluth, Rosenbluth, Teller, and Teller 1953; Tanner and Wong 1987). As such, Bayesian inference can be made via random draws

[5] One can expect that all individuals would be assigned to the ideal point model if the vector model is a nested or a special case of the ideal point model as argued by Carroll (1972).

from the posterior distribution. Next, we turn the discussion to the specification of prior distribution and issues related to the identification of these MDS models.

3.1.1 Selection of Prior Distribution

Prior selection in Bayesian Finite mixture models needs special considerations as the posterior distribution can be improper if a non-informative or reference prior is used for the membership parameter. This is due to the possibility of having empty segments[6] during the MCMC iterations as a sub-sample of size *n* may not contain any observations with the probability of $(1-p_j)^n$, where p_j is the weight of segment j (Diebolt and Robert 1994). Moreover, if an improper prior is chosen for segment j, then the resulting posterior distribution is also improper, and a finite mixture model becomes unidentifiable[7]. Several fixes to this problem have been proposed. These fixes include proper or partially proper priors (Roeder and Wasserman 1997), hierarchical priors (Escobar and West 1995), and data dependent priors (Richardson and Green 1997; Wasserman 2000). In this paper, we use a diffuse but proper prior for the segment membership parameter. Specifically, we assume a Dirichlet prior[8] for the segment membership variables γ_{s_1} and γ_{s_2} with hyper parameters $\boldsymbol{\alpha}_{\mathbf{s_1}} = (\alpha_1, ..., \alpha_{S_1})$ and $\boldsymbol{\alpha}_{\mathbf{s_2}} = (\alpha_1, ..., \alpha_{S_2})$.

---

[6] Empty segments refer to the segments to which no observation is allocated during the MCMC iterations.
[7] This can be easily verified by inspecting the trace plot as the posterior distribution would diverge to infinity if the posterior is improper.
[8] Density of Dirichlet distribution has the same functional form as the likelihood of Multinomial distribution (Robert and Casella, 1999).

That is, $p(\gamma_{s_1}) = D(\alpha_1,...,\alpha_{S_1})$, where $D(\alpha_1,...,\alpha_{S_1}) = \frac{\Gamma\left(\sum_{k=1}^{S_1}\alpha_k\right)}{\prod_{k=1}^{S_1}\Gamma(\alpha_k)}\prod_{k=1}^{S_1}\Gamma(\gamma_k^{\alpha_k-1})$, and this leads to a Dirichlet posterior with prior parameters reflected with the number of individuals assigned to each segment. When no information is available about the segment membership parameter vector $\boldsymbol{\gamma} = (\gamma_1,...,\gamma_s)$, one often wants to reflect this lack of information with vague or uninformative priors. One potential vague prior is the uniform prior (i.e., with $\alpha_1 = \cdots = \alpha_{s_1} = 1$) where all individuals have identical segment membership probability, and this prior produces a posterior distribution proportional to the likelihood function. In this paper, we set $\alpha_1 = \cdots = \alpha_{s_1} = 1$ to reflect a state of relative ignorance.

Next, we assign Beta prior with parameters a_0 and b_0 for structural heterogeneity parameter ϕ_0, where $f(\phi_0) = \frac{1}{B(a_0,b_0)}\phi_0^{a_0-1}(1-\phi_0)^{b_0-1}$ if $0 < \phi_0 < 1$ and 0 otherwise. As in the case of segment membership prior, we set $a_0 = b_0 = 1$ so that an individual has an equal probability to belong either the vector or ideal point model.

Next, we assign a univariate normal distribution with zero mean as a prior for all other parameters. That is, $p(x_{jt}) \sim N(0,\tau_1^2)$, $p(b_{s_1}) \sim N(0,\tau_2^2)$, $p(d_{s_2}) \sim N(0,\tau_3^2)$, $p(v_{s_1t}) \sim N(0,\tau_4^2)$, $p(w_{s_2t}) \sim N(0,\tau_5^2)$, and $p(c_{s_2}) \sim N(0,\tau_6^2)I(c_{s_2} < 0)$ respectively. As discussed before, the scale parameter c_{s_2} needs to be constrained to be negative to prevent anti-ideal point solutions. As such, we assume a right truncated normal prior for this

parameter, and this leads to a posterior with a right truncated normal distribution. A natural choice for hyperprior of the inverse of variance is a Gamma prior and thus we set $p(\tau_1^{-2}) \sim G(k_1, u_1)$, $p(\tau_2^{-2}) \sim G(k_2, u_2)$, $p(\tau_3^{-2}) \sim G(k_3, u_3)$, $p(\tau_4^{-2}) \sim G(k_4, u_4)$, $p(\tau_5^{-2}) \sim G(k_5, u_5)$, and $p(\tau_6^{-2}) \sim G(k_6, u_6)$ respectively where G denotes the Gamma distribution. We set $k_1 = ... = k_6 = u_1 = ... = u_6 = 0.5$ to reflect a lack of prior information.

3.1.2 Issues in Identification

Throughout the development of the preceding model, it has been assumed that both vector and ideal point models are identifiable. However, LCMDS models with larger dimensions than the number of segments are not identified (DeSarbo, Manrai, and Manrai 1994; Wedel and DeSarbo 1996). Therefore, the number of latent segments needs to be greater than or equal to the number of dimensions. Furthermore, Multidimensional Scaling Models are typically "under-identified" – these models have several solutions that render the same goodness of fit values. Thus, parameters need to be constrained to obtain unique solution, and these identification issues are often known as parameter indeterminacies (Wedel and DeSarbo 1996; Wedel and Kamakura 2000; Young 1987). These indeterminacies stem from the fact that orthogonal transformations of respective parameters (e.g., $\underline{x}$ and $\underline{w}$) do not alter the likelihood[9], thereby hindering the convergence of the MCMC. Many types of MDS models can be translated, expanded,

[9] Since it does not change distances in the ideal point model or scalar products in the vector model

rotated, or reflected without any effect on the solution (DeSarbo, Manrai, and Manrai 1994; DeSarbo and Rao 1986; DeSarbo and Rao 1984; Wedel and DeSarbo 1996; Young 1987). Here, in our case, both the vector and the ideal point model can be rotated or reflected without any change in the likelihood. For instance, the solution is invariant under an orthogonal transformation by a non-singular T×T matrix, R, such that $RR' = I$. It follows that $\det(R) = \pm 1$. If $\det(R) = 1$, then R represents an orthogonal rotation, and if $\det(R) = -1,$ then R represents a reflection (Erdem 1996). This rotational or reflection indeterminacy requires $T(T-1)/2$ constraints for parameters to be uniquely identified (DeSarbo and Rao 1984, 1986). Second, both the ideal point and the vector model have permutation indeterminacy, which is nothing more than a reordering of the dimensions. Algebraically, this is accomplished by applying a permutation matrix P , which is a T×T binary matrix with a single 1 in each row and column, to the matrix of brand coordinates such that $X^* = XP$. Third, the vector model suffers from scale indeterminacy between x_{jt} and v_{s_1t} . That is, multiplying a constant to x_{jt} such that $x_{jt}^* = x_{jt}a$ and dividing the same constant from v_{s_1t} such that $v_{s_1t}^* = v_{s_1t}/a$, would lead to the same solution. As such, T constraints need to be imposed to prevent this scale indeterminacy. For example, Chintagunta (1994) imposes restrictions such that some elements of the estimated vectors are fixed when he estimates a two-dimensional model, and that some elements in both brand locations and vectors are fixed if the number of dimensions is greater than two in order to remove both the scale and rotational indeterminacies. Alternatively, Andrews and Manrai (1999), and Sinha, Inman, Wang, and Park (2005) fix the locations of T^2 parameters of the vectors as an identity matrix, thereby removing the scale, reflection,

and rotational indeterminacies. Specifically, these authors propose that the preference vectors of the first T segments be projected onto the T dimensional common space while remaining segments' preferences (e.g., $T+1,...,S_1$) are projected onto the $(S_1 - T)$ segments' vectors. For example, with a two dimension and three segment solution, two dimensional vectors can be expressed as $\underline{v}_1 = (1,0,v_{s_1 1})'$ and $\underline{v}_2 = (0,1,v_{s_1 2})'$. By imposing orthogonality to the first t vector locations (t = 1,…,T), the rotational and reflection indeterminacies can be removed. Furthermore, the scale indeterminacy is removed by fixing the location of t-th element of the first T vectors to one.

For the ideal point model, the solution is also invariant under origin indeterminacy, which is sometimes called translational, additive or centering indeterminacy (Wedel and DeSarbo 1996). This origin indeterminacy can be viewed as a shift of the origin of the joint space. That is, this translation is accomplished by adding a constant to all of the coordinates on each dimension (Young 1987) such that $x^*_{jt} = x_{jt} + a$, and $w^*_{s_1 t} = w_{s_1 t} + a$, which requires T constraints. To cope with the origin indeterminacy, different attempts have been made (Chintagunta 1994; Elrod 1988; Elrod and Keene 1994; Erdem 1996). First, a restriction can be made to the brand location x_{jt} such that $\sum_{j=1}^{J} x_{jt} = 0$ for each dimension t (t = 1,…,T) (e.g., Erdem 1996). Alternatively, this origin indeterminacy can be removed by fixing the location(s) of one of brands at arbitrary constant(s) [10] (Chintagunta 1994; Elrod 1988; Elrod and Keene 1994). It should be noted

[10] Note that these two alternatives are equivalent (Erdem 1996).

that the scale indeterminacy is applicable only to the vector model, while the origin indeterminacy is applicable only to the ideal point model in our context. Table 3.1 shows different types of indeterminacies applied to vector and ideal point models.

Table 3.1: Types of indeterminacies applied to vector and ideal point model

Type of indeterminacy	Vector model	Ideal Point model
Rotation	Yes	Yes
Reflection	Yes	Yes
Permutation	Yes	Yes
Scale	Yes	No
Origin	No	Yes

Generally, these indeterminacies are acknowledged by subtracting corresponding degrees of freedom (i.e., number of indeterminacies) from the total number of model parameters in a Maximum Likelihood approach (DeSarbo and Rao 1986). In Bayesian analysis, however, these indeterminacy issues have to be circumvented either by the imposition of a strong or informative prior on brand coordinates (DeSarbo, Kim, and Fong 1999; DeSarbo, Kim, Wedel, and Fong 1998) or by post-processing of the result[11] (Oh and Raftery 2001). A notable exception is Bradlow and Schmittlein (2000). They assume a squared Mahalonobis distance between a search engine and an URL[12]. To

[11] Oh and Raftery (2001) proposed a Bayesian proximity model in which they employ the posterior mode of MCMC samples that minimizes the sum of squared residuals as an approximation of object coordinates.
[12] Uniform Resource Locator refers to the global address of documents and other resources on the World Wide Web.

remove the rotational and reflection indeterminacies, they impose T^2 constraints of which $(T+1)T/2$ search engine coordinates are fixed at zero, and $(T-1)T/2$ coordinates are constrained on the positive real space. Similarly, we impose constraints to remove the various types of indeterminacies in the ideal point model and the vector model respectively. Specifically, we fix T parameters at one for the first T brand coordinates and constrain the remaining T parameters of these brand coordinates on the positive real line. For instance, let $\underline{x} = (x_{11}, x_{21}, ..., x_{jt})' = (\underline{x}_c, \underline{x}_{nc})'$ where $\underline{x}_c$ are constrained and $\underline{x}_{nc}$ are unconstrained brand coordinates in a two dimensional model. Here, we use a Gamma prior for x_{11} and x_{22} so that these parameters are constrained on the positive real line while fixing x_{12} and x_{21} at one. As such, the reflection, rotational, and scale indeterminacies can be removed simultaneously. Finally, for the origin indeterminacy, we impose a restriction on the brand location x_{jt} such that $\sum_{j=1}^{J} x_{jt} = 0$ for each dimension t (t = 1,…,T).

3.2 Estimation

For T dimensions, S_1 vector model segments, and S_2 ideal point model segments, the joint posterior density function of the unknown parameters $(x_{jt}, w_{s_2 t}, v_{s_1 t}, b_{s_1}, c_{s_2}, d_{s_2}, \phi_0, \gamma_{s_1}, \gamma_{s_2})$ is:

$$
\begin{aligned}
&p(x_{jt}, w_{s_2t}, v_{s_1t}, b_{s_1}, c_{s_2}, d_{s_2}, \phi_0, \gamma_{s_1}, \gamma_{s_2} \mid \Delta_{ijr}) \propto \\
&\prod_{i=1}^{I} \left[\begin{array}{l} \left(\prod_{s_1=1}^{S_1} \prod_{r=1}^{R} \prod_{j=1}^{J} f_{v_{S_1}}(\Delta_{ijr} \mid x_{jt}, v_{s_1t}, b_{s_1}) \right)^{I(\chi_i=0, \xi_{is_1}=s_1)} \\ \times \left(\prod_{s_2=1}^{S_2} \prod_{r=1}^{R} \prod_{j=1}^{J} f_{IP_{S_2}}(\Delta_{ijr} \mid x_{jt}, w_{s_2t}, c_{s_2}, d_{s_2}) \right)^{I(\chi_i=1, \xi_{is_2}=s_2)} \end{array} \right] \\
&\times p(x_{jt} \mid \tau_1^2) p(w_{s_2t} \mid \tau_5^2) p(v_{s_1t} \mid \tau_4^2) p(b_{s_1} \mid \tau_2^2) p(c_{s_2} \mid \tau_6^2) I(c_{s_2} < 0) \\
&\times p(d_{s_2} \mid \tau_3^2) p(\phi_0) p(\gamma_{s_1}) p(\gamma_{s_2}) \\
&\times p(\tau_1^2) p(\tau_5^2) p(\tau_4^2) p(\tau_2^2) p(\tau_6^2) p(\tau_3^2).
\end{aligned}
\tag{3.8}
$$

Here, we assume prior independence among $x_{jt}, w_{s_2t}, v_{s_1t}, b_{s_1}, c_{s_2}, d_{s_2}, \phi_0, \gamma_{s_1},$ and γ_{s_2} (i.e.,

$$
\begin{aligned}
&p(x_{jt}, w_{s_2t}, v_{s_1t}, b_{s_1}, c_{s_2}, d_{s_2}, \phi_0, \gamma_{s_1}, \gamma_{s_2}) = p(x_{jt} \mid \tau_1^2) p(w_{s_2t} \mid \tau_5^2) p(v_{s_1t} \mid \tau_4^2) \\
&p(b_{s_1} \mid \tau_2^2) p(c_{s_2} \mid \tau_6^2) I(c_{s_2} < 0) p(d_{s_2} \mid \tau_3^2) p(\phi_0) p(\gamma_{s_1}) p(\gamma_{s_2})).
\end{aligned}
$$

Because of the complicated form of the posterior density function in equation 3.8, numerical integration is required to get a Bayesian estimate of the parameters. Therefore, we employ a Markov Chain Monte Carlo (MCMC) technique that first augments the data by the consumer specific latent variables[13], and iteratively generates new values for each parameter given the current values of other parameters (Gilks, Richardson, and Spiegelhalter 1996). The researcher inputs $\Delta_{ijr}, T, S_1,$ and S_2. Then, random starting values are generated for initial parameter values. Then, this MCMC iterates the following steps (A detailed description of the full conditional distributions is provided in the Appendix).

[13] This refers to consumer specific latent variables such as the structural heterogeneity indicators, as well as the vector and the ideal point model segment membership indicators.

1. Generate individual structural heterogeneity indicator $I(\chi_i = 0)$ and $I(\chi_i = 1)$.
2. Generate individual level vector segment membership indicator $I(\xi_{is_1} = s_1)$ for $s_1 = 1,...,S_1$.
3. Similarly, generate individual level ideal point segment membership indicator $I(\xi_{is_2} = s_2)$ for $s_2 = 1,...,S_2$.
4. Generate additive parameter $b_{s_1} \mid \tau_2^2, \chi_i, \xi_{is_1}, x_{jt}, v_{s_1t}$ for $s_1 = 1,...,S_1$.
5. Generate vector parameter $v_{s_1t} \mid \tau_4^2, \chi_i, \xi_{is_1}, x_{jt}, b_{s_1}$ for $s_1 = 1,...,S_1$.
6. Generate scale parameter $c_{s_2} \mid \tau_6^2, \chi_i, \xi_{is_2}, x_{jt}, w_{s_2t}, d_{s_2}$ for $s_2 = 1,...,S_2$.
7. Generate additive parameter $d_{s_2} \mid \tau_3^2, \chi_i, \xi_{is_2}, x_{jt}, w_{s_2t}, c_{s_2}$ for $s_2 = 1,...,S_2$.
8. Generate w_{s_2t} for $s_2 = 1,...,S_2$ and t = 1,…,T.
9. Generate x_{jt} for j = 1,…,J and t = 1,…,T.
10. Update hyper-prior τ_k^2, for $k = 1,...,6$.
11. Update $\phi_0 \mid \chi_i, \alpha_0, \beta_o$.
12. Update $\gamma_{s_1} \mid \alpha_{s_1}, \xi_{is_1}$.
13. Update $\gamma_{s_2} \mid \alpha_{s_2}, \xi_{is_2}$.

This process is iterated first for a designated number of burn-in cycles, and then for a pre-set number of estimation iterations. Note that for sake of convenience in interpretation of the resulting space, we have an option in the program to set $c_{s_2} = -1$ for all ideal point segments S_2 when using preference data ($c_{s_2} = 1$ for dispreference data).

3.2.1 Model Selection and Convergence Checking

Model selection for finite mixture MDS models is usually considered a question of finding the optimal number of vector model segments (s_1^*), ideal point model segments (s_2^*) and dimension t^* . In order to do this, two general approaches have been proposed. One approach is to estimate each model separately for each $s_1 (= 1, ..., S_1)$ and $s_2 (= 1, ..., S_2)$, and then select the best model by computing the posterior densities of the different specifications for the observed data for a preset dimension T assuming $S \geq T$. The other approach is to assume the number of segments and dimension T as random. Therefore, the latter approach requires to develop a method that simulates over the larger model of random $s_1 (= 1, ..., S_1)$, $s_2 (= 1, ..., S_2)$, and T, and then to find posterior probabilities for each s_1 , s_2 , and T according to the proportion of time that the simulation spends in each subset of s_1 , s_2 , and T (Green 1995; Richardson and Green 1997).

While the second approach has its own merits, it is computationally infeasible given the large number of vector model segments (S_1), ideal point model segments (S_2), and the number of dimension T for the proposed model. As such, we estimate each model separately given fixed s_1 , s_2 , and t, and compare the posterior densities of these specifications.

Let H_k denote a specification of the proposed model $(k = 1,..., S_1 + S_2)$[14]. The posterior density $\Pr(D \mid H_k)$ is obtained by integrating the posterior probability over the parameter space:

$$\Pr(D \mid H_k) = \int \Pr(D \mid \Theta_k, H_k) f(\Theta_k \mid H_k) d\Theta_k \,, \tag{3.9}$$

where $\Pr(D \mid \Theta_k, H_k)$ is the likelihood function given a specification H_k, data D, and parameters Θ_k, and $f(\Theta_k \mid H_k)$ is the prior density. Therefore, equation 3.9 is often called the marginal likelihood of the data. Sometimes it is also called the integrated likelihood, the marginal probability, or the predictive probability of the data (Kass and Raftery 1995; Newton and Raftery 1994). This marginal likelihood can also be used for model comparison through the calculation of Bayes factors. A Bayesian comparison of any two model H_i and H_j via the Bayes factor is defined as follows:

$$BF_{ij} = \frac{\Pr(H_i \mid D)}{\Pr(H_j \mid D)} = \frac{\Pr(D \mid H_i)\Pr(H_i)}{\Pr(D \mid H_j)\Pr(H_j)} \,. \tag{3.10}$$

The ratio of marginal likelihoods $BF_{ij} = \Pr(D \mid H_i) / \Pr(D \mid H_j)$ is known as the Bayes factor if $\Pr(H_i) = \Pr(H_j)$ is assumed. That is, when models are considered equally probable a priori, then the Bayes factor is equal to the posterior odds. Kass and Raftery

[14] For the sake of simplicity, we assume that we fix the dimension T.

(1995) suggested a guideline for any comparison of two model specifications displayed in Table 3.2.

Table 3.2: A Guideline for any comparison of two model specifications (Kass and Raftery 1995)

$2\log_e(BF_{ij})$	BF_{ij}	Evidence against H_j
0 to 2	1 to 3	Weak
2 to 6	3 to 20	Positive
6 to 10	20 to 150	Strong
> 10	> 150	Very Strong

As the Bayes factor requires the comparison of any two models, the logarithm of the marginal probability of the data has been often reported. This logarithm of the marginal probability leads to an interpretation of the Bayes factor that does not assume one of the specifications as the true specification (Kass and Raftery 1995).

Since this marginal likelihood is not analytically tractable, a body of literature on the use of an approximation of the marginal likelihood has been developed (Evans and Swartz 1995; Kass and Raftery 1995). Here, we provide a brief catalogue description of approximation methods.

The first method we review is the harmonic mean of the posterior likelihood:

$$\hat{p}_1(D) = \left\{ \frac{1}{(N-m)} \sum_{i=m+1}^{N} \Pr(D \mid \Theta_k^{(i)}, H_k)^{-1} \right\}^{-1}, \tag{3.11}$$

where N is the total number of draws, m is the burn-in draws, and $\Theta^{(i)}$ is the i-th random deviates from the joint posterior distribution under a specification H_k. Although it is easy to compute, the harmonic mean estimator($\hat{p}_1(D)$) does not satisfy the Gaussian Central Limit theorem, and can be unstable in some applications since it can be overly influenced by small likelihood values (Newton and Raftery 1994).

Newton and Raftery (1994) suggested estimators that employ an importance sampling function. Let $f(\Theta_k) = \delta p(\Theta_k) + (1-\delta) p(\Theta_k \mid D)$ be a mixture of the prior and posterior for parameter Θ_k of the model K where δ is small ($0 \le \delta \le 1$). Sampling from $f(\Theta_k)$ is straightforward, and the new estimate $\hat{p}_3(D)$ is defined by:

$$\hat{p}_3(D) = \frac{\sum_{i=m+1}^{N} p(D \mid \Theta_k^{(i)}, H_k) / \left\{ \delta \hat{p}_3(D) + (1-\delta) p(D \mid \Theta_k^{(i)}, H_k) \right\}}{\sum_{i=m+1}^{N} \left\{ \delta \hat{p}_3(D) + (1-\delta) p(D \mid \Theta_k^{(i)}, H_k) \right\}^{-1}}, \qquad (3.12)$$

where δ is a number between 0 and 1. It should be noted that if $\delta = 0$ then $\hat{p}_3(D)$ reduces to $\hat{p}_1(D)$. Although it satisfies the Gaussian Central Limit theorem, $\hat{p}_3(D)$ requires simulations from both the prior and the posterior distribution. To cope with this irksome aspect, Newton and Raftery (1994) proposed an estimator by simulating all (N-m) values from the posterior:

$$\hat{p}_4(D) = \frac{\delta(N-m)/(1-\delta) + \sum_{i=m+1}^{N} \Pr(D \mid \Theta_k^{(i)}, H_k) / \left\{ \delta \hat{p}_4(D) + (1-\delta) \Pr(D \mid \Theta_k^{(i)}, H_k) \right\}}{\delta(N-m)/(1-\delta) \hat{p}_4^{-1}(D) + \sum_{i=m+1}^{N} \left\{ \delta p_4(D) + (1-\delta) \Pr(D \mid \Theta_k^{(i)}, H_k) \right\}^{-1}}. \qquad (3.13)$$

This estimator $\hat{p}_4(D)$ can be easily evaluated by using an iterative scheme: first, insert an initial value on the right hand side for $\hat{p}_4(D)$ and calculate a new $\hat{p}_4(D)$. Next, this new value of $\hat{p}_4(D)$ goes into the right hand side iteratively until the difference between two consecutive values of $\hat{p}_4(D)$ becomes negligible.

Although $\hat{p}_1(D)$ and $\hat{p}_4(D)$ are easy to implement, they often tend to prefer more complex models (Lopes 2000). Lopes (2000) showed that both $\hat{p}_1(D)$ and $\hat{p}_4(D)$ often favor higher dimensional models than the true model structure. Secondly, these estimators can be unstable by the low likelihood values. Finally, the estimate may not be stable depending on the specification of δ.

Like $\hat{p}_4(D)$, if a researcher wants to avoid the introduction of the prior distribution for the approximation of the marginal likelihood, the Schwarz Criterion or BIC (Bayesian Information Criteria)[15] can be a rough approximation to the logarithm of the Bayes factor, and provide a reasonable indication of the highest posterior probability (Kass and Raftery 1995). The BIC criterion with the highest posterior probability is calculated as:

$$BIC = -2\Pr(D \mid \hat{\Theta}_k, H_k) + \log(N)NP, \tag{3.14}$$

[15] Schwarz Criterion is calculated as $\Pr(D \mid \hat{\Theta}_k, H_k) - \frac{1}{2}\log(N)NP$

where N denotes the number of observations; $\hat{\Theta}_k$ is the Maximum Likelihood Estimator (MLE) under H_k. LL denotes the log likelihood; and NP denotes the number of parameters. Here, $\hat{\Theta}_k$ is chosen to be one of the simulated parameter estimates that result in the highest log likelihood. Rust et al. (1995) compared various model selection heuristics and concluded that the BIC or Schwarz Criterion is the most consistently accurate model selection criteria (Rust, Simester, Brodie, and Nilikant 1995).

Finally, we use another model selection criterion similar to the Reversible Jump MCMC methods (RJMCMC). RJMCMC methods treat a specification H_k as unknown over $k \in K$ and are useful for exploring posterior distributions for model parameters in the context of uncertainty of a model H_k (Green 1995). A feasible approach for doing RJMCMC is to use a "mini" RJMCMC as described by Lopes (2000). The basic premise of this approach is to use the posterior or likelihood at each iteration for each model $k = 1,...,K$. In addition to model specific priors specified above, we also need to specify the marginal "Model" prior probability, $p(k)$, over $k \in K$. This algorithm can be described as follows:

Step 1. Propose a new visit to a new model k' according to transition probability $p(k' \mid k) = J(k \to k')$. We assume that we would have K equally likely

models and the probability to move from a model *k* to model *k'* to be uniform (i.e., $p(k'|k) = \frac{I(k' \neq k)}{K-1}$)[16].

Step 2. Propose a new candidate $\Theta_k^{(n)}$ from the posterior distribution $p(\Theta_k | D)$ and $\Theta_{k'}^{(n)}$ from $p(\Theta_{k'} | D)$ and set $\Theta_j^{(n)} = \Theta_j$ for all $j \neq k, k'$.

Step 3. Accept the new model *k'* with probability α :

$$\alpha = \min\left\{1, \left(\frac{p(k|k')p(D|\Theta_k)p(\Theta_k)p(D|\Theta_{k'})p(\Theta_{k'})}{p(k'|k)p(D|\Theta_k^{(n)})p(\Theta_k^{(n)})p(D|\Theta_{k'}^{(n)})p(\Theta_{k'}^{(n)})} \times \frac{p(D|\Theta_{k'}^{(n)})p(\Theta_{k'}^{(n)})p(\Theta_k^{(n)})\prod_{j\neq k,k'} p(\Theta_j)p(k')}{p(D|\Theta_k)p(\Theta_k)p(\Theta_{k'})\prod_{j\neq k,k'} p(\Theta_j)p(k)} \right)\right\} \quad (3.15)$$

$$= \min\left\{1, \frac{p(D|\Theta_{k'})p(k')}{p(D|\Theta_k)p(k)}\right\}.$$

In the following chapters, we present the results of a simulated data and an application to the prescription behavior of medical doctors in the antidepressant medication category. Here, we report BIC, $\hat{p}_1(D)$, $\hat{p}_4(D)$, and RJMCMC as model selection criteria.

---

[16] $\frac{p(k'|k)}{p(k|k')}$ reduces to 1.

Chapter 4

A SIMULATION STUDY

4.1 Introduction

In Chapter 2, we contend that LCMDS models need to account for structural heterogeneity, and that failure to account for both structural heterogeneity and preference heterogeneity can lead to a misrepresentation of true market structure in a product category. As such, we propose an approach that explicitly accommodates mixtures of vector and ideal point representation as a form of structural heterogeneity, and that accommodates preference heterogeneity via a finite mixture modeling. In order to illustrate this point more clearly, we present the results from a simulation in which the true market structure and two types of heterogeneity are created and thus known for hypothesized consumers. The three main purposes of the simulation are: (a) to demonstrate the proposed model's ability to uncover the true preference structure and correctly account for both types of heterogeneity; (b) to understand whether model selection criteria based on derived solutions point to the true structure; and (c) to demonstrate that failure to account for both structural and preference heterogeneity may lead to the misrepresentation of the true underlying consumer's preference structure.

Synthetic data was generated for S = 5 segments in T = 2 dimensions. To reflect structural heterogeneity, we created 2 vector segments ($S_1 = 2$) and 3 ideal point segments ($S_2 = 3$). As discussed in Chapter 3, we set $c_{s_2} = -1$ for simplicity and

convenience. Once the coordinates of parameters (i.e., brands, vectors, ideal points, and additive constant parameters) were randomly generated, we added random errors[17] to the derived utility for each consumer to simulate within-segment heterogeneity. This synthetic data is comprised of I = 180 consumers whose preferences towards J = 5 brands are generated in R = 6 replications. Figure 4.1 shows the created joint space configuration underlying this synthetic data in two dimensions. One can easily identify preference order of each segment from Figure 4.1. For instance, Vector Segment 1 prefers Brand D the most and Brand E the least, while Vector Segment 2 prefers Brand E the most. Similarly, Ideal Point Segment 1 shows the highest preference toward Brand B and the lowest preference toward Brand E.

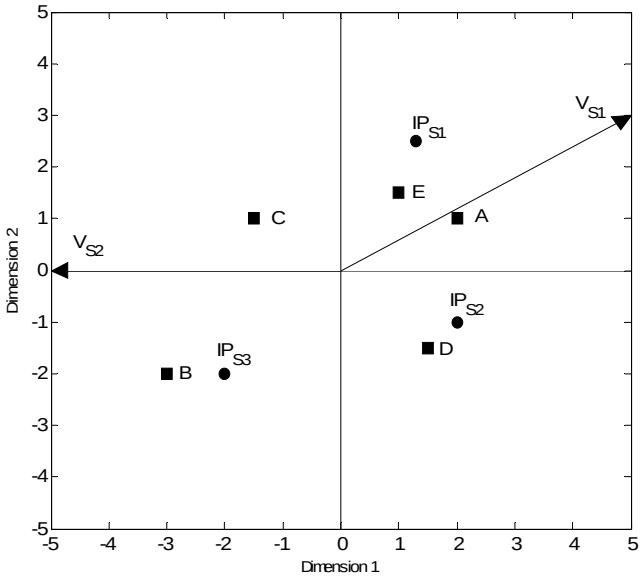

Figure 4.1: Original configuration underlying synthetic data

[17] Random errors refer to random numbers generated from the standard normal distribution.

We then calibrated the proposed model with the first 5 replications per individual, and use the 6th replication for model validation purposes. As discussed in Chapter 3, we used an incremental search strategy to find the best fitting model.

4.2 Results

We ran the MCMC for 10,000 iterations. The first 5,000 iterations were used as burn-in, and the last 5,000 iterations were used for generating model estimates. Convergence was checked by starting the chain from multiple starting points and by inspecting time series plots. In addition to the plot assessment, we used a standard CODA function that provides convergence diagnostics (Cowles and Carlin 1996).

The rest of this chapter is organized as follows: first, we discuss the model selection criteria. Next, we discuss the recovery of the true parameters. Finally, we compare models with structural heterogeneity and models without structural heterogeneity.

We first examine various model selection criteria discussed in Chapter 3. Table 4.1 displays the results of model selection criteria for t = 1, 2, 3 and S = 1,…,12. It should be noted that models with T > S are not identified and denoted by N/A in Table 4.1 (Wedel and DeSarbo 1996). As shown, all model selection criteria clearly point to the true structure (i.e., 2 dimensional, 2 segment vector and 3 segment ideal point model). For example, the best fitting model shows harmonic mean = -6368, Newton and Raftery estimator = -6381, and BIC = 12931. Moreover, the RJMCMC method shows the

probability of 1 for the true structure to be selected[18]. Table 4.1 also displays several interesting patterns: (a) models with structural heterogeneity show higher harmonic mean and Newton and Raftery estimator, and lower BIC than models without structural heterogeneity give the same number of segments. For example, the 2 dimensional, 5 segment ideal point model shows harmonic mean = -8270, Newton and Raftery estimator = -8279, and BIC = 16736, and the 2 dimensional, 5 segment vector model shows harmonic mean = -11097, Newton and Raftery estimator = -11151, and BIC = 22390; (b) if the number of segments is higher than that of the true structure, all model selection criteria tend to get worse. For instance, the 2 dimensional, 2 segment vector and 4 segment ideal point model has harmonic mean = -8095. This phenomenon is also observed in models without structural heterogeneity. That is, the 2 dimensional, 6 segment ideal point model results in harmonic mean = -8305, which is clearly lower than that of the 2 dimensional, 5 segment ideal point model (-8270); and (c) the BIC shows relatively consistent and accurate results despite its simple computation.

[18] Not shown in Table 4.1

Table 4.1: Comparison of model selection criteria on simulated data

		Dimension 1			Dimension 2			Dimension 3		
Vector Segment	Ideal Point Segment	Harmonic Mean	Newton and Raftery	BIC	Harmonic Mean	Newton and Raftery	BIC	Harmonic Mean	Newton and Raftery	BIC
0	1	-348617	-348620	697288	N/A*	N/A	N/A	N/A	N/A	N/A
0	2	-271648	-271655	543365	-111569	-111598	223257	N/A	N/A	N/A
0	3	-268357	-268553	536802	-71577	-71743	143298	-69139	-69308	138482
0	4	-268310	-268367	536723	-34728	-34915	69627	-29418	-29706	59074
0	5	-268306	-268338	536732	-8270	-8279	16736	-6661	-7011	13593
0	6	-268307	-268340	536749	-8305	-8564	16831	-7241	-7560	14787
1	0	-325526	-325530	651105	N/A	N/A	N/A	N/A	N/A	N/A
1	1	-268532	-268537	537133	-111575	-111652	223270	N/A	N/A	N/A
1	2	-146029	-146062	292144	-71557	-71583	143259	-173856	-173904	347915
1	3	-146027	-146105	292156	-29140	-29272	58450	-29042	-29214	58322
1	4	-146025	-146095	292170	-7173	-7336	14541	-29489	-29850	59248
1	5	-146185	-146230	292507	-7474	-7497	15169	-7670	-8096	15644
1	6	-146210	-146266	292573	-7477	-7606	15199	-6851	-7117	14040
2	0	-115717	-115722	231504	-111573	-111642	223265	N/A	N/A	N/A
2	1	-168502	-168508	337091	-136341	-136347	272826	-136340	-136368	272884
2	2	-165210	-165217	330523	-49227	-49237	98624	-29347	-29503	58932
2	3	-165161	-165176	330441	**-6368**	**-6381**	**12931**	-29569	-29828	59409
2	4	-165161	-165178	330458	-8095	-8261	16410	-48920	-48958	98145
2	5	-165158	-165177	330470	-29009	-29027	58264	-34875	-35077	70088
2	6	-165159	-165179	330488	-8175	-8677	16621	-52042	-52743	104456
3	0	-76903	-76907	153891	-77770	-77837	155684	-71496	-71580	143197
3	1	-68612	-68619	137327	-29892	-29911	59953	-36115	-36814	72467
3	2	-68611	-68634	137342	-7245	-7263	14686	-6556	-6870	13383
3	3	-68611	-68620	137358	-10157	-10858	20535	-23300	-24001	46905
3	4	-68609	-68624	137370	-8419	-8884	17084	-49205	-49904	98749
3	5	-68609	-68616	137388	-6580	-6594	13431	-49839	-50538	100050
3	6	-68608	-68623	137402	-6580	-6599	13457	-7324	-8025	15055
4	0	-69675	-69679	139453	-54320	-54412	108809	-50917	-51039	102071
4	1	-61713	-61717	123545	-8948	-8966	18092	-6431	-6779	13134
4	2	-38806	-38812	77749	-8949	-8966	18118	-6569	-6627	13442
4	3	-165161	-165178	330475	-11071	-11493	22388	-6582	-6617	13501
4	4	-61710	-61733	123589	-8155	-8202	16581	-6380	-6452	13132
4	5	-61710	-61718	123606	-8179	-8250	16654	-6567	-6615	13540
4	6	-38806	-38813	77817	-8948	-8960	18217	-6579	-6602	13597
5	0	-31635	-31642	63389	-11097	-11151	22390	-6647	-6804	13565
5	1	-31635	-31641	63406	-8949	-8964	18118	-6586	-6617	13476
5	2	-68610	-68616	137373	-10983	-11446	22211	-6571	-6597	13480
5	3	-68608	-68614	137386	-8157	-8192	16585	-7755	-8454	15881
5	4	-68608	-68615	137403	-11103	-11182	22502	-7602	-8301	15609
5	5	-31635	-31641	63473	-8948	-8958	18218	-6697	-7395	13835
5	6	-31635	-31641	63490	-8217	-8283	16781	-7844	-8543	16160
6	0	-31635	-31640	63406	-11102	-11185	22424	-6632	-6963	13568
6	1	-31635	-31641	63423	-11100	-11179	22445	-6641	-6840	13620
6	2	-31635	-31640	63440	-11103	-11194	22478	-7632	-8331	15636
6	3	-69107	-69146	138402	-8948	-8956	18193	-6635	-6769	13675
6	4	-38806	-38812	77817	-11101	-11165	22523	-6588	-6621	13615
6	5	-31635	-31641	63490	-8151	-8852	16648	-7553	-8252	15579
6	6	-38806	-38814	77850	-10937	-11179	22244	-7726	-8425	15958

*N/A: Models with T > S are not estimable

Next, we discuss the parameter recovery of the proposed model. The time series plots for the 2 segment vector and 3 segment ideal point model over the last 5,000 iterations of MCMC are displayed in Figure 4.2. Figure 4.3 displays the marginal posterior densities of the parameters in terms of histograms of the posterior samples. As shown in Figures 4.2 and 4.3, we do not observe any evidence of label switching commonly reported in the Bayesian finite mixture literature, although we do not impose any restrictions to prevent the label switching (Stephens 2000). Histograms in Figure 4.3 show that there is no multiple mode of marginal posterior of the estimated parameters.

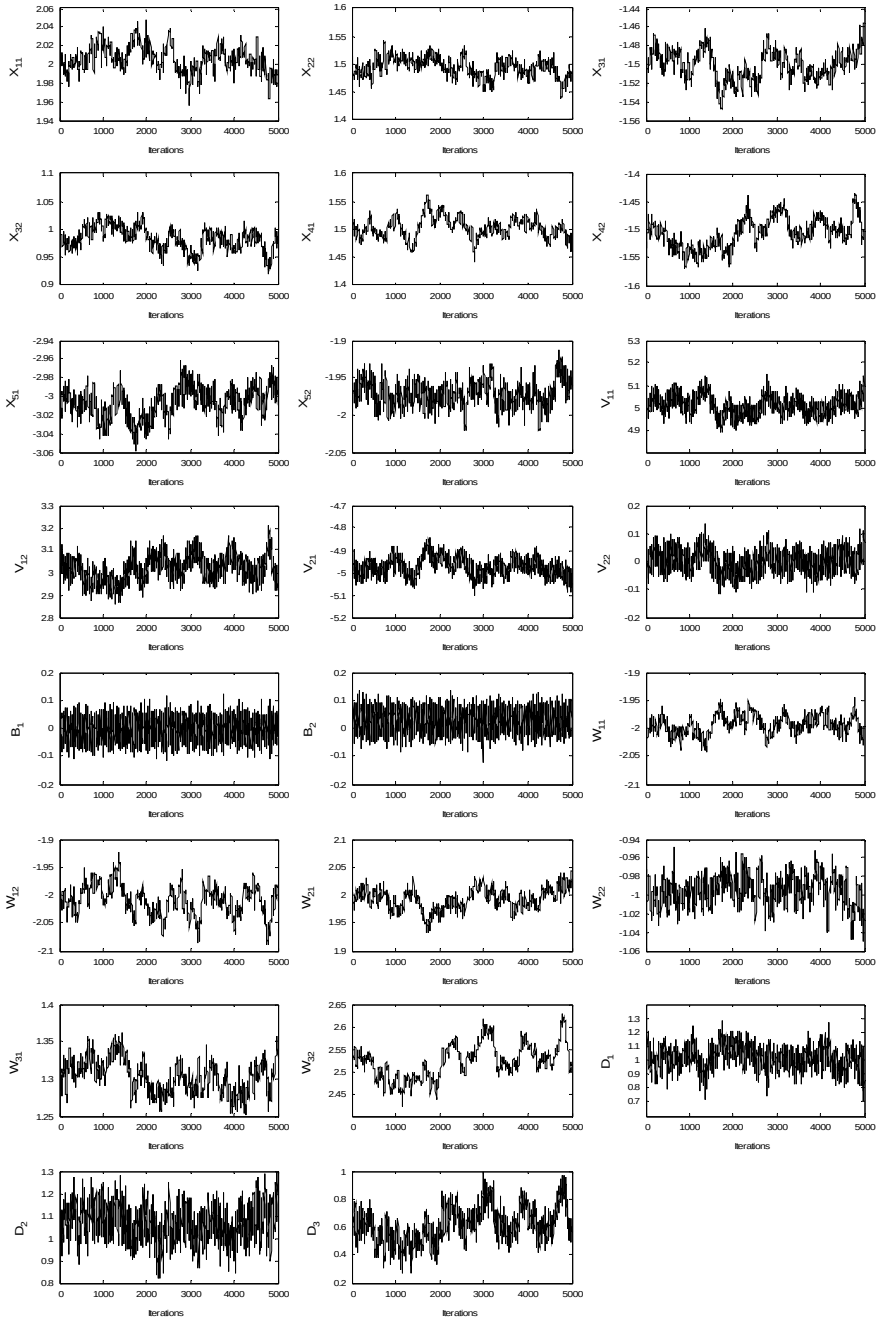

Figure 4.2: Time-series plots of the parameters over the last 5,000 iterations of MCMC

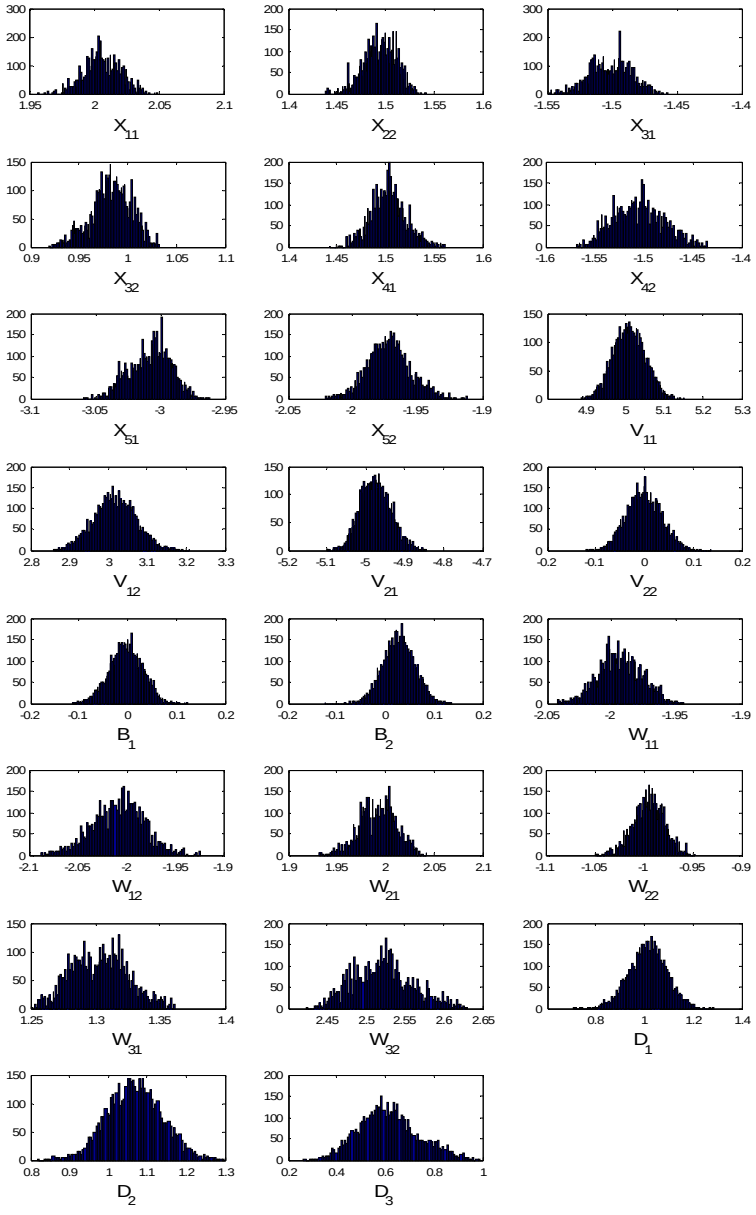

Figure 4.3: Histogram of marginal posterior of the estimated parameters for the 2 dimensional, 2 segment vector and 3 segment ideal point model

Point estimates of parameters and their posterior standard deviations are displayed in Table 4.2 along with true parameters. Note that two brand coordinates are fixed at 1 throughout MCMC iterations for the identification purpose, and that posterior standard deviations are reported in parentheses.

Table 4.2: Comparison of True Parameters and Estimates

	Simulated True		Parameter Estimates		RMSE
	Dimension 1	Dimension 2	Dimension 1	Dimension 2	
Brands					0.0329
A	2.00	1.00	2.005 (0.014)	1.00*	
B	1.00	1.50	1.00*	1.495 (0.017)	
C	-1.50	1.00	-1.501 (0.016)	0.983 (0.020)	
D	1.50	-1.50	1.503 (0.018)	-1.506 (0.026)	
E	-3.00	-2.00	-3.007 (0.016)	-1.972 (0.017)	
Ideal Points					0.0365
Segment 1	1.30	2.50	1.303 (0.022)	2.522 (0.038)	
Segment 2	2.00	-1.00	1.993 (0.018)	-0.995 (0.016)	
Segment 3	-2.00	-2.00	-1.994 (0.017)	-2.010 (0.027)	
Ideal Point Additive Constants					0.1365
Segment 1	0.52		0.615 (0.122)		
Segment 2	1.12		1.068 (0.071)		
Segment 3	1.09		1.014 (0.076)		
Vectors					0.0734
Segment 1	5.00	3.00	5.011 (0.041)	3.017 (0.053)	
Segment 2	-5.00	0.00	-4.975 (0.039)	-0.001 (0.036)	
Vector Additive Constants					0.0277
Segment 1	0.00		-0.003 (0.034)		
Segment 2	0.00		0.028 (0.033)		
Average RMSE					0.0296

* Fixed at 1 throughout MCMC iterations
Note: Values inside Parentheses are Posterior Standard Deviations

The estimation results in Table 4.2 show that the proposed model uncovers the true parameters fairly well with small standard deviations. Next, we compare the RMSE (Root Mean Squared Error) between the true parameters and the estimates. The last

column of Table 4.2 shows RMSE between the true parameters and the estimated parameters. The average RMSE across all model parameters is less than 3% and the parameter specific RMSEs show that the proposed model recovers the true structure of the synthetic data.

We then compare the best fitting 2 dimensional models (i.e., 5 segment vector model, 5 segment ideal point model, and 2 segment vector and 3 segment ideal point model) in terms of derived 2 dimensional space as shown in Figure 4.4. Earlier, we postulated that failure to accommodate both structural and preference heterogeneity may result in misrepresentation of consumer's preference structure and a biased market structure. First, the derived map of the 2 dimensional, 5 segment vector model (panel b) shows that brand coordinates are almost collapsed into 1 dimension (e.g., brand coordinates are almost linearly positioned) and somewhat exaggerates the degree of competition between Brand A and Brand B. Furthermore, Brands C and D are aligned on the second dimension. This suggests that Brands C and D differentiate only in terms of the first dimension in two dimensional space. On the contrary, the derived map of the 2 dimensional, 5 segment ideal point model (panel c) shows that the coordinates of brands are more widely dispersed than those of the true structure, and that the two segments (i.e., Segment 3 and Segment 5) are hardly distinguished from the derived space.

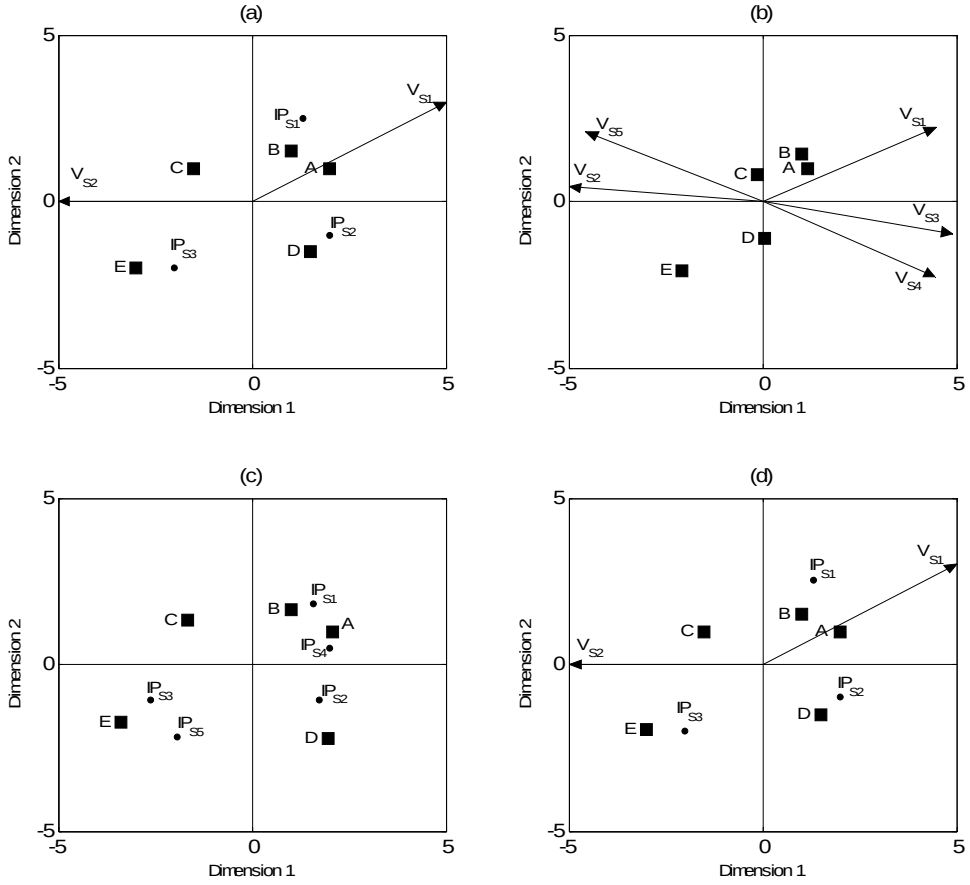

Figure 4.4: Comparisons derived and original MDS plots for (a) true structure, (b) 2 dimensional, 5 segment vector model (the lengths of vectors are resized to facilitate interpretation), (c) 2 dimensional, 5 segment ideal point model, and (d) 2 dimensional, 2 segment vector and 3 segment ideal point model

Note: Vs_1: Vector Segment 1, IPs_1: Ideal Point Segment 1

Next, we compare RMSE measures for several models. First, we compare aggregated solutions-models with T = 1 and S = 1 - with the true structure. Table 4.3 shows RMSE measures of various models for both calibration sample and holdout sample. The RMSE of the aggregated models for the calibration data are 26.72 for the 1 dimensional, 1 segment vector model and 27.67 for the 1 dimensional, 1 segment ideal point model while the best fitting 2 dimensional, 2 segment vector and 3 segment ideal point model exhibits 2.23 of RMSE. The performance of the model that accounts for structural and preference heterogeneity is 12 times better than those of models that do not account for any type of heterogeneity. A similar pattern can be found in holdout sample: 1 dimensional vector model has RMSE = 11.98 and 1 dimensional ideal point model shows RMSE = 12.40. Consistent with the calibration data, these aggregated models show approximately 12 times worse performance than that of the best fitting model.

Next, we compare RMSE of the best fitting model with structural heterogeneity and the best fitting models without structural heterogeneity. As shown in Table 4.3, the model with structural heterogeneity shows lower RMSE both in the calibration sample and in the holdout sample.

Table 4.3: RMSE (Root Mean Squared Error) for various models

Dimension	Vector Segment	Ideal Point Segment	Calibration Sample	Holdout Sample
1	1	0	26.72	11.98
1	0	1	27.67	12.40
2	2	3	2.23	1.04
2	5	0	3.93	1.78
2	0	5	3.03	1.36

Together, the results from this synthetic data suggest that failure to properly account for both structural and preference heterogeneity produces structures that either misrepresent or are greatly inaccurate of the true underlying preference structure. It should be noted that further extensive Monte Carlo testing is required in future research where many synthetic data sets are examined as a number of data, error, and model specification forms are experimentally manipulated.

Chapter 5

EMPIRICAL APPLICATION

5.1 Introduction

A major US pharmaceutical company conducted a market research study among physicians (i.e., general practitioners, psychiatrists, etc.) in order to understand their prescribing decisions for antidepressant prescription medications. Antidepressants are prescribed for symptoms such as depression, social anxiety disorder, and generalized anxiety disorder (GAD). Symptoms of depression include: (1) a sad feeling that will not go away; (2) restlessness or slowed movements; (3) changes in appetite or weight; (4) changes in sleeping patterns; (5) fatigue or lack of energy; (6) feeling worthless or feeling guilty for no reason; or (7) repeated thoughts of death or suicide[19]. To be diagnosed as having major depression, a person must show at least five of the above symptoms. People with social anxiety disorder have an extreme, constant fear of one or more social or public situations. GAD is characterized by feelings of excessive anxiety and worry that cannot be controlled and are present for at least 6 months. Treatments of depression include: (1) antidepressant medication; (2) a variety of psychotherapeutic approaches; (3) electroconvulsive therapy (ECT); and (4) other treatments (e.g., light therapy).

---

[19] Depression Guide Panel (2006)

Currently, the depression therapy market is one of the largest medication markets in the world[20]**.** It is estimated that approximately 6% of U.S. population – some 19 million people – will have a depressive illness that warrants treatment[21]. In 2004, global sales of branded antidepressants exceeded $14 billion, and U.S. sales totaled $9.9 billion.

The data is comprised of 250 doctors' prescriptions for five leading brands of antidepressants and the number of prescriptions for each brand recorded on a monthly basis over a 7 month period. Due to confidentiality agreements with the client pharmaceutical company, the specific brands are anonymously labeled with letters A to E. The leading five brands in the data can be categorized into three types of inhibitors based on their chemical components: (1) SSRI (Selective serotonin reuptake inhibitor); (2) SNRI (Serotonin and norepinephrine reuptake inhibitor); and (3) NDRI (Norepinephrine and dopamine reuptake inhibitor). SSRIs are known to increase the brain's level of serotonin, thereby improving mood, and are particularly helpful in heading off depression in the early stages. Three brands in our data such as Brand B, Brand C, and Brand D belong to this category. SNRIs are believed to work especially well for the people (up to 40 percent) who don't respond to serotonin-related antidepressants (or SSRIs). As such, an SNRI is usually prescribed as a second-line medication. Brand A is an SNRI. Lastly, NDRI is a selective catecholamine (norepinephrine and dopamine) reuptake inhibitor, and it has only a small effect on serotonin reuptake. Brand E is an NDRI. These five brands can also be classified by their

---

[20] Mental Health Business Week (2006)
[21] Consumer Reports Best Buy Drugs (2005)

intended treatment. Table 5.1 shows the indication of treatment for various symptoms by brands (Psychiatric 2000).

Table 5.1: Intended treatment of various symptoms

	Obsessive-Compulsive Disorder	Post-Traumatic Stress Disorder	Premenstrual Dysphoric Disorder	Body Dysmorphic Disorder	Generalized Anxiety Disorder	Social Anxiety Disorder	Depression
Brand A	0	0	0	0	1	1	1
Brand B	0	0	0	0	1	0	1
Brand C	1	1	0	0	1	1	1
Brand D	1	1	1	0	0	1	1
Brand E	0	0	0	0	0	0	1

Note: 1 refers when a brand can be applied to the particular symptom

Side effects may occur in a number of patients taking any medication, and are typically dependent on dosage and blood level. Many side effects are more likely to occur at the initiation of treatment or within a short time following dosage increases, and patients often adapt to side effects over time. Some common side effects are headache, nausea, diarrhea, dizziness, sweating, tremor, and dry mouth. These common side effects are relatively minor and usually go away in time (or are short-lived). However, there are some side effects that are not minor and may become bothersome or sometimes dangerous. These side effects include: nervousness and agitation, feeling of panic or dread, increased thoughts of suicide, insomnia, drowsiness or confusion, loss of libido or difficulty of achieving erections, and weight gain.

Brand specific characteristics merit further explanation. First, Brand B was introduced recently compared to other brands. Brand B also has a relatively small likelihood of side effects, as this medication can be given in small doses, and is known to cause less interaction with other drugs compared to other SSRIs[22]. Compared to other brands, Brand D is the only brand approved for obsessive-compulsive disorder (OCD) in children and adolescents age 6-17 years. Although it can be prescribed to the variety of symptoms, Brand C is strongly warned against female patients who are pregnant given its possible teratogenic effect. Brand E has the lowest incidence of sexual dysfunction. Consumer Report (2006) shows that out of 1664 patients, Brand E has the lowest sexual dysfunction side effect. However, if this medication is taken in increased dosage (e.g., 450 mg/day), Brand E has a higher risk of seizure. Also, Brand E is associated with the development of psychotic symptoms, including delusions and hallucinations, and is recommended to use cautiously in patients with psychotic disorders. As discussed earlier, Brand A is believed to work especially well for the patients who don't respond to SSRIs, and is usually recommended as a second-line medication. It also has the lowest elimination half-life[23] (5 hours vs. 24 hours of other antidepressants).

Antidepressants need to be taken for at least 4 to 8 weeks to become beneficial. Although much progress has been made in developing medications for treating depression, the exact causes and optimal treatments of depression have not been resolved (Berndt, Cockburn, and Griliches 1996). The American Psychiatric Association issued

---

[22] Brand A and Brand D are also known to have few drug interaction (Antidepressant Comparison Chart, www.RxFiles.ca)

[23] The elimination half-life of a drug refers to the time course necessary for the quantity of the xenobiotic agent in the body to be reduced to half of its original level through various elimination processes.

medical practice guidelines for the prescription of antidepressants that include: (1) anticipated side effects and their safety or tolerability; (2) history of prior response in patient or family member; (3) patient preference; (4) cost; and (5) quantity and quality of clinical trial data.

In addition, the American Psychiatric Association suggests the following recommendations regarding various type of concurrent general medical disorders: (1) SSRIs (Brands B, C, and D) and Brand E are safer choices if patients have cardiac disease; (2) Brands B, C, and D have less anticholinergic activity[24]. Therefore, these brands are recommended if patients have glaucoma; (3) dose-dependent elevations in blood pressure with Brand E are usually mild, but more severe elevations have been observed. As such, Brand E should be less preferable in patients with hypertension; and (4) Brand E shows a beneficial effect on the symptoms of Parkinson's disease.

Doctors usually recommend an antidepressant that is least likely to cause side effects for the person taking it. A recent study led by National Institute of Mental Health, however, shows that less than 30% of patients who take their first-line medication have significant remission (Menza 2006). As such, if the patient shows no response or partial response to the medication, doctors usually change dose, switch to other antidepressants, or add a second antidepressant medication from a different class (American Psychiatric Association 2000, p.17). This type of prescription behavior is reflected in the current data. It should be noted that there is no further information whether each prescription represents a new prescription, renewal, or mixing prescriptions of multiple brands.

---

[24] Anticholinergic activity refers side effects such as dry mouth, blurred vision, urinary hesitancy, or constipation

Existing literature (e.g., Manchanda and Chintagunta 2004) has focused on impact of detailing effort on the prescription behavior of individual physicians. This study is somewhat limited by the lack of such marketing vehicle information.

The data is comprised of a total of 162,515 prescriptions over a seven month period. Prescription share is: Brand A (22.2%), Brand B (21.3%), Brand C (16.3%), Brand D (16.4%), and Brand E (23.9%) respectively. Table 5.2 shows descriptive statistics on the number of prescriptions by brands. On average, a physician in our sample writes approximately 2.65 prescriptions for an antidepressant per month and 18 prescriptions over 7 month periods. This data, however, shows fairly large heterogeneity for the number of prescriptions.

Table 5.2: Descriptive statistics on the number of prescription

	Mean	Standard Deviation	Maximum	Minimum
Brand A	21	17	99	0
Brand B	20	17	103	0
Brand C	15	11	88	0
Brand D	15	13	94	0
Brand E	22	13	97	0

Note, in the pharmaceutical industry, practitioners usually focus on the prescription volume for their detailing efforts. As such, practitioners tend to pay more attention to high volume physicians (Manchanda, Rossi, and Chintagunta 2004).

As our data is count data, we transform this data by double mean centering in order to estimate the proposed model. This double-mean centering can somehow minimize the predominance of rows (i.e., prescription volume) and column (i.e., brand share) effects in the resulting spatial solution. Double mean centering is performed as follows:

$$\Delta^{*}_{ijr} = \Delta_{ijr} - \frac{1}{IR}\sum_{i=1}^{I}\sum_{r=1}^{R}\Delta_{ijr} - \frac{1}{J}\sum_{j=1}^{J}\Delta_{ijr} + \frac{1}{IJR}\sum_{i=1}^{I}\sum_{r=1}^{R}\sum_{j=1}^{J}\Delta_{ijr}\,, \qquad (5.1)$$

where Δ_{ijr} is the number of prescriptions of a doctor i for brand j in occasion r, and Δ^{*}_{ijr} denotes the mean centered data. Figure 5.1 compares distributions of the original prescription data and double mean centered data.

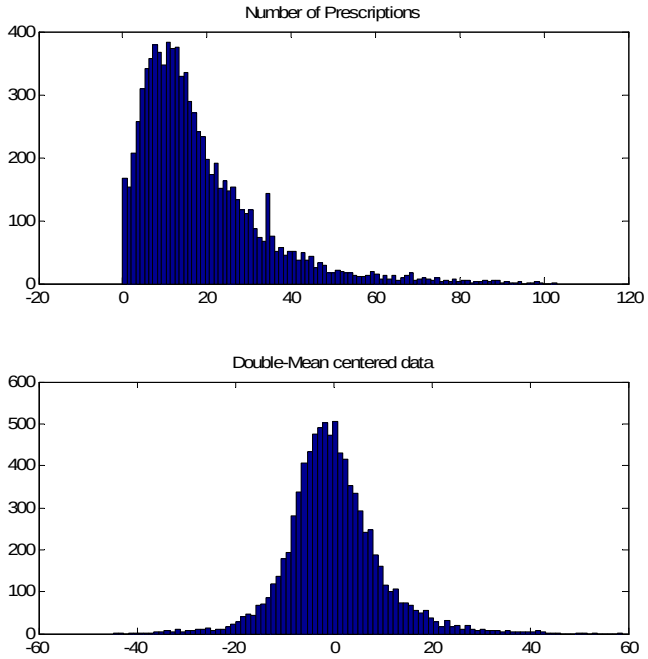

Figure 5.1: Distribution of double mean centered data

We then calibrate the proposed model based on the first six months' prescriptions (140,404 prescriptions), and use the last month's prescriptions for model validation (22,111 prescriptions). Note that we set $c_{s_2} = -1$ similar to the analysis of the simulated data to facilitate the interpretation of the derived MDS map. The rest of this chapter is organized as follows: we first present the model selection heuristics for the proposed model over different dimensions and number of segments. We then report the model validation results. Next, we discuss the best fitting solution of the proposed model. Finally, we discuss the marketing implications of the results.

5.2 Empirical Results

As discussed in Chapter 3 and Chapter 4, we use the incremental search strategy to find the best fitting model. In this strategy, we estimate each model separately by increasing the number of the vector segment s_1 ($s_1 = 1,...,S_1$), the number of the ideal point segment s_2 ($s_2 = 1,...,S_2$), and dimension t ($t = 1,...,T$), and select the best fitting model by comparing various model selection criteria. Results of model selection criteria as the Harmonic mean $\hat{p}_1(D)$, Newton and Raftery's estimator $\hat{p}_4(D)$, BIC, and RJMCMC are presented in 5.3**.** Results in Table 5.3 suggest that the two dimensional model is most parsimonious for this data. The two dimensional, four segment vector and four segment ideal point model shows the highest harmonic mean (-182750) and Newton and Raftery's estimator (-183090), and lowest BIC (365780). As a two dimensional

model is selected, we restrict ourselves for the incremental search to two dimensional models without structural heterogeneity. As shown in Table 5.3, model selection criteria point that the two dimensional, ten segment ideal point model and the two dimensional, ten segment vector model as the most parsimonious model among models without structural heterogeneity. However, models without structural heterogeneity show lower harmonic mean, Newton and Raftery's estimator, and higher BIC than the best fitting model with structural heterogeneity. For instance, two dimensional, ten segment vector model has 1650 lower harmonic mean, 2006 lower Newton and Raftery's estimator, and 3360 higher BIC than the two dimensional four segment vector and four segment ideal point model. Likewise, two dimensional, ten segment ideal point model has 1405 lower harmonic mean, 1763 lower Newton and Raftery's estimator, and 2870 higher BIC than the two dimensional four segment vector and four segment ideal point model.

In addition to model selection criteria discussed so far, we compare the model selection probability using Reversible Jump MCMC method discussed in Chapter 3. Here, we ran 10 million iterations with likelihoods of models estimated so far in order to get marginal posterior probability from the RJMCMC method, and used the last 5 million iterations to calculate the model selection probability. As shown in Tables 5.3 and 5.4, the Reversible Jump MCMC method (RJMCMC) also points the two dimensional, four segment vector and four segment ideal point model with the highest probability to be chosen (prob. = 1).

Table 5.3: Comparison of model selection criteria

		Dimension 1				Dimension 2				Dimension 3			
Vector Segment	Ideal Point Segment	Harmonic Mean	Newton and Raftery	BIC	RJ-MCMC	Harmonic Mean	Newton and Raftery	BIC	RJ-MCMC	Harmonic Mean	Newton and Raftery	BIC	RJ-MCMC
0	1	-359665	-359710	719387	0.0000	N/A*	N/A	N/A	N/A	N/A	N/A	N/A	N/A
0	2	-359665	-359708	719405	0.0000	-278954	-279267	558036	0.0000	N/A	N/A	N/A	N/A
0	3	-359665	-359716	719423	0.0000	-254339	-254840	508833	0.0000	-250513	-251214	501243	0.0000
0	4	-359665	-359705	719439	0.0000	-245024	-245587	490230	0.0000	-244917	-245617	490085	0.0000
0	5	-359666	-359720	719459	0.0000	-221521	-222221	443250	0.0000	-249939	-250640	500166	0.0000
1	0	-359238	-359270	718531	0.0000	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
1	1	-320018	-320089	640109	0.0000	-278404	-278456	556935	0.0000	N/A	N/A	N/A	N/A
1	2	-319805	-319871	639701	0.0000	-250658	-250820	501470	0.0000	-246740	-247293	493696	0.0000
1	3	-319814	-319909	639737	0.0000	-227791	-228482	455762	0.0000	-224676	-224691	449605	0.0000
1	4	-319807	-319913	639740	0.0000	-212391	-212951	424990	0.0000	-210187	-210306	420662	0.0000
1	5	-319811	-319899	639766	0.0000	-206366	-207060	412967	0.0000	-194867	-195091	390058	0.0000
2	0	-279405	-279577	558882	0.0000	-278421	-278458	556968	0.0000	N/A	N/A	N/A	N/A
2	1	-255606	-255699	511303	0.0000	-244447	-244704	489048	0.0000	-244030	-244036	488275	0.0000
2	2	-255605	-255699	511319	0.0000	-229685	-230029	459551	0.0000	-226347	-226638	452945	0.0000
2	3	-255605	-255672	511336	0.0000	-209300	-209310	418807	0.0000	-211084	-211153	422455	0.0000
2	4	-255604	-255701	511352	0.0000	-201228	-201928	402691	0.0000	-197013	-197249	394348	0.0000
2	5	-255604	-255702	511371	0.0000	-194192	-194892	388645	0.0000	-185721	-185980	371801	0.0000
3	0	-256585	-256918	513260	0.0000	-251261	-251438	502676	0.0000	-245247	-245097	490710	0.0000
3	1	-242777	-242894	485663	0.0000	-223674	-223988	447529	0.0000	-224090	-224207	448432	0.0000
3	2	-246744	-246847	493615	0.0000	-209258	-209889	418723	0.0000	-215191	-215484	430670	0.0000
3	3	-242780	-242893	485705	0.0000	-198894	-199587	398023	0.0000	-209871	-210372	420065	0.0000
3	4	-246745	-246850	493653	0.0000	-186830	-186884	373921	0.0000	-186033	-186734	372424	0.0000
3	5	-246715	-246720	493611	0.0000	-188491	65535	377270	0.0000	-186156	-186849	372705	0.0000
4	0	-245663	-246120	491435	0.0000	-229898	-230078	459977	0.0000	-225370	-225447	450991	0.0000
4	1	-239359	-239469	478845	0.0000	-210232	-210551	420672	0.0000	-208293	-208591	416874	0.0000
4	2	-233954	-233959	468052	0.0000	-207215	-207499	414664	0.0000	-200191	-200432	400704	0.0000
4	3	-239359	-239475	478881	0.0000	-186725	-186833	373710	0.0000	-187794	-188175	375946	0.0000
4	4	-243871	-243981	487924	0.0000	**-182747**	**-182915**	**365781**	**1.0000**	-183468	-184160	367329	0.0000
4	5	-233999	-234124	468197	0.0000	-184707	-185409	369730	0.0000	-186712	-187297	373855	0.0000
5	0	-239892	-240412	479910	0.0000	-215393	-215786	430993	0.0000	-208327	-209029	416943	0.0000
5	1	-230427	-230583	460999	0.0000	-197473	-197545	395180	0.0000	-194817	-194919	389957	0.0000
5	2	-230449	-230625	461060	0.0000	-188213	-188231	376687	0.0000	-188723	-189422	377804	0.0000
5	3	-231122	-231239	462425	0.0000	-187453	-187460	375193	0.0000	-187522	-188110	375438	0.0000
5	4	-231123	-231235	462444	0.0000	-184617	-184657	369548	0.0000	-184566	-185259	369563	0.0000
5	5	-242974	-243109	486164	0.0000	-183635	-184336	367611	0.0000	-187080	-187779	374626	0.0000

* N/A: Model is not estimable

Table 5.4: Model selection criteria for two dimensional models without structural heterogeneity

Vector Segment	Ideal Point Segment	Harmonic Mean	Newton and Raftery	BIC	RJ-MCMC
6	0	-207990	-208680	410746	0.0000
7	0	-201321	-202016	402901	0.0000
8	0	-191851	-192547	383989	0.0000
9	0	-187685	-188380	375684	0.0000
10	**0**	**-184400**	**-185096**	**369141**	**0.0000**
11	0	-184784	-185485	369935	0.0000
0	6	-207985	-208684	416205	0.0000
0	7	-207122	-207821	414505	0.0000
0	8	-202899	-203598	406086	0.0000
0	9	-187560	-188257	375435	0.0000
0	**10**	**-184155**	**-184853**	**368652**	**0.0000**
0	11	-187517	-188214	375402	0.0000

Thus, a comparison of these model selection heuristics shows that the two dimensional, four segment vector and four segment ideal point model is the most parsimonious model for this data. All model selection criteria show that the models with structural heterogeneity outperform models without structural heterogeneity. This result is consistent with the findings of existing literature that incorporate structural heterogeneity (Gilbride and Allenby 2004; Jedidi and Kohli 2005).

In addition to these model selection criteria, we also calculated VAF (Variance Accounted For) and RMSE (Root Mean Squared Error) for both calibration and hold out sample. Table 5.5 shows the result of these validation measures.

Table 5.5: VAF and RMSE

			VAF		RMSE	
Dimension	Vector Segment	Ideal Point Segment	Calibration	Holdout	Calibration	Holdout
1	0	1	0.000	0.000	23.758	9.729
1	1	0	0.001	0.043	23.744	9.804
2	0	8	0.454	0.378	4.343	4.931
2	8	0	0.477	0.371	4.337	4.765
2	0	10	0.501	0.414	6.875	7.446
2	10	0	0.498	0.395	6.878	7.579
2	4	4	0.503	0.385	6.842	7.665

Similar results can be seen in VAF (Variance Accounted For) and RMSE (Root Mean Squared Error) for both calibration and holdout sample as shown in Table 5.5. The best fitting two dimensional, four segment vector and four segment ideal point model shows huge improvement over aggregated models (i.e., 1 dimensional 1 segment vector or ideal point model). The model with both structural and preference heterogeneity improves the model fit by 250% over the aggregated model. For the same number of segments, this two dimensional, four segment vector and four segment ideal point model outperforms models without structural heterogeneity both in calibration and holdout sample.

Next, we devote the remainder of this section to discussing the result of the two dimensional, four segment vector and four segment ideal point model. The estimates and the respective posterior standard deviations of the estimates are reported in Table 5.6. These parameters can be better understood visually, as is commonly done in most Multidimensional Scaling methods. Figure 5.2 shows the derived joint space for the two

dimensional, four segment vector and four segment ideal point model in two dimensional space.

Table 5.6: Parameter estimates for the two dimensional, four segment vector and four segment ideal point model

	Dimension 1	Dimension 2	Overall
Brands			
A	-0.412 (0.020)	-2.805 (0.022)	
B	-1.911 (0.016)	-1.150 (0.014)	
C	1.865 (0.018)	1.00*	
D	-0.542 (0.020)	1.264 (0.018)	
E	1.00*	1.690 (0.014)	
Ideal Points			
Segment 1	1.256 (0.025)	0.247 (0.024)	
Segment 2	-1.015 (0.024)	0.392 (0.021)	
Segment 3	1.513 (0.022)	-1.113 (0.019)	
Segment 4	0.012 (0.019)	-2.396 (0.020)	
Ideal Point Model Additive Constants			
Segment 1			6.205 (0.073)
Segment 2			5.740 (0.072)
Segment 3			8.082 (0.097)
Segment 4			10.153 (0.136)
Vectors			
Segment 1	13.216 (0.186)	-3.606 (0.141)	
Segment 2	3.283 (0.080)	5.835 (0.071)	
Segment 3	-2.706 (0.080)	-10.686 (0.100)	
Segment 4	-12.714 (0.175)	5.129 (0.138)	
Vector Model Additive Constants			
Segment 1			0.000 (0.081)
Segment 2			-0.001 (0.074)
Segment 3			0.000 (0.063)
Segment 4			0.000 (0.055)

* Fixed at 1 throughout MCMC iterations

Note: Values inside parentheses are posterior standard deviations

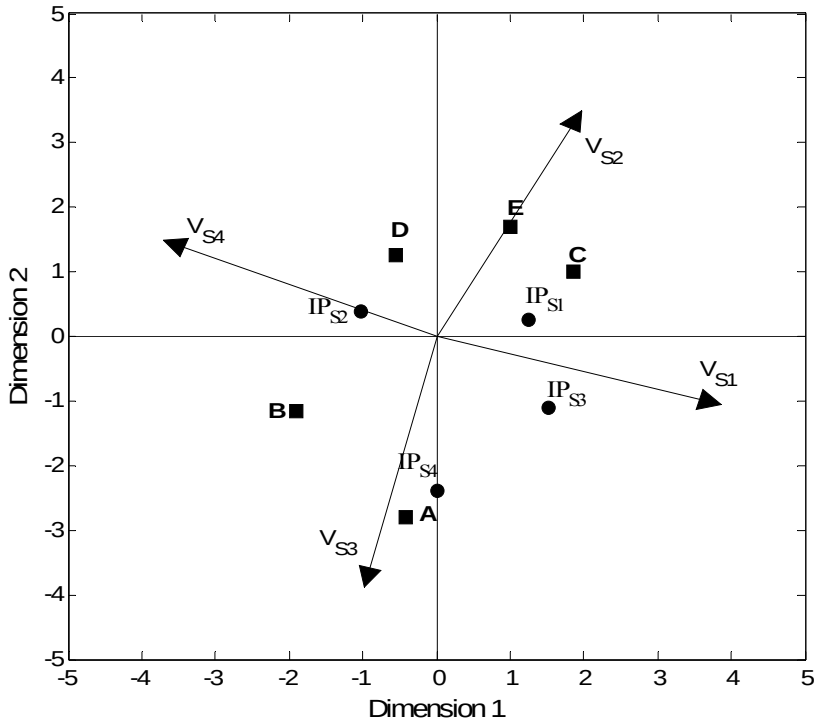

Figure 5.2: Derived joint space for the two dimensional, four segment vector and four segment ideal point model

Note: the lengths of vectors are normalized to constant length while retaining their orientations to facilitate the interpretation of derived joint map

The derived joint space map shows that there are three distinct groups of brands: a group comprised of Brand C, D, and E, while Brand B and Brand A are uniquely positioned. This makes sense since Brand B is recently introduced and only the medication that can be prescribed to obsessive-compulsive disorder (OCD) in children and adolescents, and Brand A is usually prescribed as a second-line medication.

Next, one can identify that dimension 1 is highly related to interaction with other drugs and psychiatric disorder associated side effects. In other words, Brands A, B, and D have less drug interaction effects than Brands C or E. Also, Brand D can safely be prescribed to children or adolescents age 6-17 years. Compared to these brands in the left side quadrant, Brand C has the highest possibility to cause teratogenic effects to female patients who are pregnant, and it has pharmacokinetic interaction with antiepileptic drugs[25] (Balabanov and Kanner 2004). Brand E can cause seizures if the dosage is too high (e.g., higher than 450 mg/day). In the other hand, the brands in the left side quadrant (i.e., Brands A, B, and D) seem to be highly correlated to psychiatric disorder related side effects such as Nervousness Restlessness ($r = -.902$, $p = 0.036$) and Increased Appetite ($r = -.892$, $p = 0.042$). This dimension is also correlated with the year of FDA approval as Brand B is the last medication and Brand E is the first medication approved by FDA.

Dimension 2 can be best explained as whether a particular brand is a first-line or second-line medication. Brand A is often used as a second-line medication for patients who have treatment-resistant depression (TRD) and has 10% advantage in remission rates when compared to SSRIs (Thase, Friedman, and Howland 2000). However, this Brand A is somehow correlated with some anticholinergic side effects, especially nausea. The dimension 2 is also correlated with such side effects as edema[26] ($r = -0.819$, $p = 0.09$) and pelvic pain[27] ($r = -0.903$, $p = 0.036$).

---

[25] Drugs used in prevention of the occurrence of epileptic seizures

[26] Swelling of any organ or tissue due to accumulation of excess fluid

[27] This can be caused by 1) uterine, vaginal, ovarian, cervical, or cystic abnormality or damage and 2) menstrual cycle

As shown in Figure 5.2, there are eight segments (i.e., 4 vector segments and 4 ideal point segments) identified by the proposed model. Ideal Point Segment 3 is the largest segment (28%), followed by Ideal Point Segment 1 (26%), Ideal Point Segment 2 (21%), and Ideal Point Segment 4 (14%), while vector segments comprise of very small number of physicians. The biggest vector segment is Vector Segment 4 (4%), followed by Vector Segment 3 (3%), and both Vector Segment 1 (2%) and Vector Segment 2 (2%) are the smallest segments.

Preferences for the ideal point segments can be inferred by the distance between brands' coordinates and segments' ideal point coordinates. For instance, Ideal Point Segment 1 shows the highest preference toward Brand C, and the lowest preference toward Brand A. Similarly, Ideal Point Segment 2 exhibits the highest preference toward Brand D while the lowest preference toward Brand A. Preferences for the vector segments can be inferred from the orthogonal projection onto the segment's vector for each brand coordinate. From Figure 5.2, one can infer that Vector Segment 1 would have the highest preference toward Brand C and the lowest preference toward Brand B. Similarly, Brand E is the most preferred brand while Brand A is the least preferred brand for Vector Segment 2, and Brand A is the most preferred brand for Vector Segment 3.

Table 5.7 presents recovered preference structure by the best fitting two dimensional, four segment vector and four segment ideal point model in terms of the mean centered data and the average number of prescriptions. This table shows two interesting aspects.

Table 5.7**:** Average Preference towards Brands and Number of Prescriptions

	Brand A	Brand B	Brand C	Brand D	Brand E	Average Number of Prescription	Std.Dev*
Vector Segment 1	4.7	-21.1	21.0	-11.7	7.1	34.1	16.3
Vector Segment 2	-17.7	-13.0	12.0	5.6	13.1	28.6	14.6
Vector Segment 3	31.1	17.5	-15.7	-12.0	-20.8	42.7	24.3
Vector Segment 4	-9.1	18.4	-18.6	13.4	-4.0	36.3	16.1
Ideal Point Segment 1	-5.9	-5.8	5.3	1.9	4.1	13.6	5.1
Ideal Point Segment 2	-4.8	2.6	-2.9	4.8	0.0	17.0	4.7
Ideal Point Segment 3	1.5	-3.6	3.5	-1.8	0.0	13.7	3.8
Ideal Point Segment 4	9.8	4.9	-4.8	-3.6	-7.5	25.7	9.1

* Std.Dev refers to standard deviation of the number of prescriptions across brands

First, preference order of a vector segment is similar to that of an ideal point segment. For instance, Vector Segment 2 has preference order of E > C > D > B > A. This pattern can be seen in Ideal Point Segment 1. Similarly, both Vector Segment 3 and Ideal Point Segment 4 show the highest preference toward Brand A and the least preference toward Brand E. Vector Segment 1 and Ideal Point Segment 3 show similar preference order. This implies that there is little difference in terms of the overall preference structure between vector segments and ideal point segments. However, we also find that vector segments can be distinguished from ideal point segments by the following: vector segments show higher preference variance across brands than ideal point segments as shown in the last two columns in Table 5.7. Preference variance across brands measured by standard deviation at the segment level shows that vector segments exhibit at least three times as high standard deviation as ideal point segments. This implies that vector segments show more heterogeneous preferences structure across

brands than ideal point segments. For instance, while both Vector Segment 1 and Ideal Point Segment 3 prefer Brand C most, the average preference score for Vector Segment 1 is 7 times higher than that of Ideal Point Segment 3. This implies that any change of product attribute can induce great change in preference as vector utility is formulized as a bilinear multiplicative form.

However, the most striking difference between vector segments and ideal point segments is prescription volume. As discussed earlier, physicians with high volume prescription draw greater attention from pharmaceutical company for detailing effort (Manchanda, Rossi, and Chintagunta 2004). Table 5.8 illustrates prescription share across segments per brand.

Table 5.8: Prescription Share Across Segments per Brand

	Brand A	Brand B	Brand C	Brand D	Brand E	Segment Size
Vector Segment 1	4%	1%	7%	3%	4%	2%
Vector Segment 2	2%	2%	5%	5%	5%	2%
Vector Segment 3	12%	10%	5%	5%	4%	3%
Vector Segment 4	6%	13%	4%	13%	7%	4%
Ideal Point Segment 1	12%	12%	25%	19%	24%	26%
Ideal Point Segment 2	14%	22%	13%	24%	19%	21%
Ideal Point Segment 3	24%	17%	26%	15%	22%	28%
Ideal Point Segment 4	26%	23%	15%	17%	14%	14%

Although Vector Segment 3 comprises only 3% of physicians, its prescription share for Brands A and B is equivalent to that of Ideal Point Segment 1 that is comprised of 26% of all physicians in our data. Vector Segment 4 also shows substantial

prescription share for Brands B and D. Manchanda, Rossi, and Chintagunta (2004) found that highly detailed physicians have low marginal responses to detailing, yet they are considered “opinion leaders” and targeted heavily. Our result provides the preference structure of physicians at the segment level. With this detailed segment level preference structure and prescription volume information, pharmaceutical firms can efficiently target physicians for their detailing effort.

Chapter 6

DISCUSSION

6.1 Contributions of the Thesis

Over the past two decades, numerous Latent Class Multidimensional Scaling (LCMDS) methods have been developed, and these methods have been used as one of the primary vehicles for simultaneous segmentation of consumers and product positioning. Two distinct types of MDS models (e.g., the vector and ideal point model) have been used to represent consumers' preferences. However, little attention has been given to structural heterogeneity for such spatial models in contrast to the large body of literature on preference heterogeneity in LCMDS models. We have focused on the possibility that sample of consumers may exhibit heterogeneous utility formulation that can be explained better by mixtures of the vector and the ideal point model. We then have introduced a new LCMDS model that explicitly incorporates both structural heterogeneity and preference heterogeneity in a generalized framework. Specifically, we model mixtures of the vector and the ideal point model to represent structural heterogeneity, and accommodated preference heterogeneity via a finite mixture method. Thereby, our approach enables one to identify unknown market segments that exhibit heterogeneous perceptions toward brands.

We then applied the proposed approach to simulated data and an actual application concerning doctors' prescriptions of five brands of antidepressant

medications. The results demonstrate that a model which incorporates both structural and preference heterogeneity outperforms models without structural heterogeneity. This finding is quite consistent with existing literature (Jedidi and Kohli 2005, Gilbride and Allenby 2004). Our approach produces segment level estimates of preference structure via vector and ideal point representations.

6.2 Directions for Future Research

Opportunities for future research merit discussion. The focus of the proposed model has been to represent consumers' preferences on derived joint spatial space via a Bayesian finite mixture approach. Accommodation of a random effect model via additive intercepts may improve the model fit and better explain preference heterogeneity. Also, we have assumed that consumers' preferences remain constant across time, and that competition among brands also remains constant across time (i.e., a stationary joint space). Researchers have shown that these assumptions are sometimes violated (DeSarbo, Fong, Liechty, and Coupland 2005; Liechty, Fong, and DeSarbo 2005). These assumptions could be relaxed in future work. We also have assumed that a set of consumers would formulate their preferences via the vector model utility while other would follow the ideal point model utility. However, a consumer may exhibit vector model utility in one occasion while this consumer may show ideal point model utility in another occasion. As such, incorporation of occasional type structural heterogeneity would be valuable. We have assumed that all consumers would share the same

dimensionality in joint space map. A number of researchers have shown that consumers may not process all available information, but only limited information to make decisions (Bettman, Luce, and Payne 1998). Future research might incorporate this behavioral process by applying a Bayesian variable selection procedure to the set of attributes.

Finally, more extensive Monte Carlo simulation analyses with synthetic data of known structure should be conducted to examine parameter recovery and structure heterogeneity as a number of data, degree of error, and model specification forms are varied.

Appendix

Markov Chain Monte Carlo Step for the Proposed Model

Estimation of the model parameters proceeds by recursively sampling from the following full conditional distributions:

1. **Generate the individual structural heterogeneity indicator $I(\chi_i = 0)$ and $I(\chi_i = 1)$.**

   $I(\chi_i = 0) \sim Binomial\big(1, p(\chi_i = 0)\big)$

   $I(\chi_i = 1) = 1 - I(\chi_i = 0)$

   where the structural heterogeneity probability $p(\chi_i = 0)$ is computed by:

$$p(\chi_i = 0 \mid \sim) = \frac{\prod_{r=1}^{R}\prod_{j=1}^{J}\left(\sum_{s_1=1}^{S_1} f_{v_{s_1}}(\Delta_{ijr} \mid x_{jt}, b_{s_1}, v_{s_1 t}) I(\xi_{i1} = s_1)\right) p(\chi_i = 0)}{\prod_{r=1}^{R}\prod_{j=1}^{J}\left(\sum_{s_1=1}^{S_1} f_{v_{s_1}}(\Delta_{ijr} \mid x_{jt}, b_{s_1}, v_{s_1 t}) I(\xi_{i1} = s_1)\right) p(\chi_i = 0) + \prod_{r=1}^{R}\prod_{j=1}^{J}\left(\sum_{s_2=1}^{S_2} f_{IP_{s_2}}(\Delta_{ijr} \mid x_{jt}, w_{s_2 t}, c_{s_2}, d_{s_2}) I(\xi_{i2} = s_2)\right) p(\chi_i = 1)}$$

$$= \frac{\prod_{r=1}^{R}\prod_{j=1}^{J}\left(\sum_{s_1=1}^{S_1} f_{v_{s_1}}(\Delta_{ijr} \mid x_{jt}, b_{s_1}, v_{s_1 t}) I(\xi_{i1} = s_1)\right) \phi_0}{\prod_{r=1}^{R}\prod_{j=1}^{J}\left(\sum_{s_1=1}^{S_1} f_{v_{s_1}}(\Delta_{ijr} \mid x_{jt}, b_{s_1}, v_{s_1 t}) I(\xi_{i1} = s_1)\right) \phi_0 + \prod_{r=1}^{R}\prod_{j=1}^{J}\left(\sum_{s_2=1}^{S_2} f_{IP_{s_2}}(\Delta_{ijr} \mid x_{jt}, w_{s_2 t}, c_{s_2}, d_{s_2}) I(\xi_{i2} = s_2)\right) \phi_1}$$

$$= \frac{L_{i|s_1}\phi_0}{L_{i|s_1}\phi_0 + L_{i|s_2}(1-\phi_0)},$$

   and $p(\chi_i = 1) = 1 - p(\chi_i = 0)$. Here, $p(\chi_i = 0 \mid \sim)$ indicates the full conditional distribution of χ_i, given all other parameters.

2. Generate the individual level vector segment membership indicator

$I(\xi_{is_1} = s_1)$ **for** $s_1 = 1,...,S_1$**.**

$$I(\xi_{is_1} = s_1) \sim Multinomial\left(p(\xi_{is_1} = 1),..., p(\xi_{is_1} = S_1)\right),$$

where the vector model segment membership probability, $p(\xi_{is_1} = s_1)$ is computed by:

$$p(\xi_{is_1} = s_1 \mid \sim) = \begin{cases} \dfrac{\prod_{r=1}^{R}\prod_{j=1}^{J} f_{v_{s_1}}(\Delta_{ijr} \mid x_{jt}, v_{s_1 t}, b_{s_1})\gamma_{s_1}}{\sum_{k_1=1}^{K_1}\prod_{r=1}^{R}\prod_{j=1}^{J} f_{v_{k_1}}(\Delta_{ijr} \mid x_{jt}, v_{k_1 t}, b_{k_1})\gamma_{k_1}} & if \quad I(\chi_i = 0) \\ \gamma_{s_1} & if \quad I(\chi_i = 1) \end{cases}$$

Similarly, generate the individual level ideal point segment membership indicator $I(\xi_{is_2} = s_2)$ for $s_2 = 1,...,S_2$.

$$I(\xi_{is_2} = s_2) \sim Multinomial\left(p(\xi_{is_2} = 1),..., p(\xi_{is_2} = S_2)\right),$$

where the ideal point model segment membership probability, $p(\xi_{is_2} = s_2)$ is computed by:

$$p(\xi_{is_2} = s_2 \mid \sim) = \begin{cases} \dfrac{\prod_{r=1}^{R}\prod_{j=1}^{J} f_{IP_{s_2}}(\Delta_{ijr} \mid x_{jt}, w_{s_2 t}, c_{s_2}, d_{s_2})\gamma_{s_2}}{\sum_{k_2=1}^{K_2}\prod_{r=1}^{R}\prod_{j=1}^{J} f_{IP_{k_2}}(\Delta_{ijr} \mid x_{jt}, w_{k_2 t}, c_{k_2}, d_{k_2})\gamma_{k_2}} & if \quad I(\chi_i = 1) \\ \gamma_{s_2} & if \quad I(\chi_i = 0) \end{cases}$$

3. Generate the additive parameter $b_{s_1} \mid \tau_2^2, \chi_i, \xi_{is_1}, x_{jt}, v_{s_1t}$ for $s_1 = 1,...,S_1$.

$$
\begin{aligned}
p(b_{s_1} \mid \sim) &\propto \prod_{i=1}^{I}\left(\sum_{s_1=1}^{S_1}\left(\prod_{r=1}^{R}\prod_{j=1}^{J} f_{v_{s_1}}(\Delta_{ijr} \mid x_{jt}, v_{s_1t})\right) I(\xi_{is_1} = s_1)\right)^{I(\chi_i=0)} p(b_{s_1}) \\
&\propto \prod_{i=1}^{I}\prod_{s_1=1}^{S_1}\prod_{r=1}^{R}\prod_{j=1}^{J} f_{v_{s_1}}(\Delta_{ijr} \mid x_{jt}, v_{s_1t})^{I(\chi_i=0)I(\xi_{is_1}=s_1)} p(b_{s_1}) \\
&\propto \prod_{i=1}^{I}\prod_{s_1=1}^{S_1}\prod_{r=1}^{R}\prod_{j=1}^{J} \exp\left(-\frac{1}{2}(\Delta_{ijr} - \sum_{t=1}^{T} x_{jt} v_{s_1t} - b_{s_1})^2\right)^{I(\chi_i=0)I(\xi_{is_1}=s_1)} \exp(-\frac{1}{2\tau_2^2} b_{s_1}^2) \\
&\propto \exp\left(-\frac{1}{2}\sum_{i=1}^{I}\sum_{s_1=1}^{S_1}\sum_{r=1}^{R}\sum_{j=1}^{J}(\Delta_{ijr} - \sum_{t=1}^{T} x_{jt} v_{s_1t} - b_{s_1})^2 I(\chi_i = 0, \xi_{is_1} = s_1) - \frac{1}{2\tau_2^2} b_{s_1}^2\right) \\
&\propto \exp\left(-\frac{1}{2}\left\{\left[\sum_{i=1}^{I}\sum_{s_1=1}^{S_1}\sum_{r=1}^{R}\sum_{j=1}^{J} I(\chi_i = 0, \xi_{is_1} = s_1) + \frac{1}{\tau_2^2}\right] b_{s_1}^2 \right.\right. \\
&\qquad \left.\left. -2\left[\sum_{i=1}^{I}\sum_{s_1=1}^{S_1}\sum_{r=1}^{R}\sum_{j=1}^{J}\left(\Delta_{ijr} - \sum_{t=1}^{T} x_{jt} v_{s_1t}\right) I(\chi_i = 0, \xi_{is_1} = s_1)\right] b_{s_1} \right\}\right) \\
&\sim N(\bar{b}, V_b^{-1}),
\end{aligned}
$$

where:

$$
\begin{aligned}
V_b &= \sum_{i=1}^{I}\sum_{s_1=1}^{S_1}\sum_{r=1}^{R}\sum_{j=1}^{J} I(\chi_i = 0, \xi_{is_1} = s_1) + \frac{1}{\tau_2^2} \\
\bar{b} &= \left(\sum_{i=1}^{I}\sum_{s_1=1}^{S_1}\sum_{r=1}^{R}\sum_{j=1}^{J}\left(\Delta_{ijr} - \sum_{t=1}^{T} x_{jt} v_{s_1t}\right) I(\chi_i = 0, \xi_{is_1} = s_1)\right) V_b^{-1}.
\end{aligned}
$$

4. Generate the vector parameter $v_{s_1t} \mid \tau_4^2, \chi_i, \xi_{is_1}, x_{jt}, b_{s_1}$ for $s_1 = 1,...,S_1$.

$$
\begin{aligned}
p(v_{s_1t} \mid \sim) &\propto \prod_{i=1}^{I}\left(\sum_{s_1=1}^{S_1}\left(\prod_{r=1}^{R}\prod_{j=1}^{J} f_{v_{s_1}}(\Delta_{ijr} \mid x_{jt}, b_{s_1})\right) I(\xi_{is_1} = s_1)\right)^{I(\chi_i=0)} p(v_{s_1t}) \\
&\propto \prod_{i=1}^{I}\prod_{s_1=1}^{S_1}\prod_{r=1}^{R}\prod_{j=1}^{J} f_{v_{s_1}}(\Delta_{ijr} \mid x_{jt}, b_{s_1})^{I(\chi_i=0)I(\xi_{is_1}=s_1)} p(v_{s_1t})
\end{aligned}
$$

$$\propto \prod_{i=1}^{I}\prod_{s_1=1}^{S_1}\prod_{r=1}^{R}\prod_{j=1}^{J}\exp\left(-\frac{1}{2}(\Delta_{ijr}-\sum_{t=1}^{T}x_{jt}v_{s_1t}-b_{s_1})^2\right)^{I(\chi_i=0)I(\xi_{is_1}=s_1)}\exp(-\frac{1}{2\tau_4^2}v_{s_1t}^2)$$

$$\propto \exp\left(-\frac{1}{2}\sum_{i=1}^{I}\sum_{s_1=1}^{S_1}\sum_{r=1}^{R}\sum_{j=1}^{J}(\Delta_{ijr}-\sum_{t=1}^{T}x_{jt}v_{s_1t}-b_{s_1})^2 I(\chi_i=0,\xi_{is_1}=s_1)-\frac{1}{2\tau_4^2}v_{s_1t}^2\right)$$

$$\propto \exp\left(-\frac{1}{2}\left\{\begin{array}{l}\left[\sum_{i=1}^{I}\sum_{s_1=1}^{S_1}\sum_{r=1}^{R}\sum_{j=1}^{J}x_{jt}^2 I(\chi_i=0,\xi_{is_1}=s_1)+\frac{1}{\tau_4^2}\right]v_{s_1t}^2 \\ -2\left[\sum_{i=1}^{I}\sum_{s_1=1}^{S_1}\sum_{r=1}^{R}\sum_{j=1}^{J}x_{jt}\left(\Delta_{ijr}-\sum_{l\neq t}^{T}x_{jl}v_{s_1l}-b_{s_1}\right)I(\chi_i=0,\xi_{is_1}=s_1)\right]v_{s_1t}\end{array}\right\}\right)$$

$$\sim N(\bar{v},V_v^{-1}),$$

where:

$$V_v=\sum_{i=1}^{I}\sum_{s_1=1}^{S_1}\sum_{r=1}^{R}\sum_{j=1}^{J}x_{jt}^2 I(\chi_i=0,\xi_{is_1}=s_1)+\frac{1}{\tau_4^2}$$

$$\bar{v}=\left(\sum_{i=1}^{I}\sum_{s_1=1}^{S_1}\sum_{r=1}^{R}\sum_{j=1}^{J}\left(\Delta_{ijr}-\sum_{l\neq t}^{T}x_{jl}v_{s_1l}-b_{s_1}\right)x_{jt}I(\chi_i=0,\xi_{is_1}=s_1)\right)V_v^{-1}.$$

5. Generate the scale parameter $c_{s_2}\mid\tau_6^2,\chi_i,\xi_{is_2},x_{jt},w_{s_2t},d_{s_2}$ for $s_2=1,...,S_2$.

$$p(c_{s_2}\mid\sim)\propto\prod_{i=1}^{I}\left(\sum_{s_2=1}^{S_2}\left(\prod_{r=1}^{R}\prod_{j=1}^{J}f_{IP_{s_2}}(\Delta_{ijr}\mid x_{jt},w_{s_2t},d_{s_2})\right)I(\xi_{is_2}=s_2)\right)^{I(\chi_i=1)}p(c_{s_2})I(c_{s_2}<0)$$

$$\propto\prod_{i=1}^{I}\prod_{s_2=1}^{S_2}\prod_{r=1}^{R}\prod_{j=1}^{J}f_{IP_{s_2}}(\Delta_{ijr}\mid x_{jt},w_{s_2t},d_{s_2})^{I(\chi_i=1)I(\xi_{is_2}=s_2)}p(c_{s_2})I(c_{s_2}<0)$$

$$\propto\prod_{i=1}^{I}\prod_{s_2=1}^{S_2}\prod_{r=1}^{R}\prod_{j=1}^{J}\exp\left(-\frac{1}{2}(\Delta_{ijr}-c_{s_2}\sum_{t=1}^{T}(x_{jt}-w_{s_2t})^2-d_{s_2})^2\right)^{I(\chi_i=1)I(\xi_{is_2}=s_2)}\exp(-\frac{1}{2\tau_6^2}c_{s_2}^2)I(c_{s_2}<0)$$

$$\propto \exp\left(-\frac{1}{2}\sum_{i=1}^{I}\sum_{s_2=1}^{S_2}\sum_{r=1}^{R}\sum_{j=1}^{J}(\Delta_{ijr}-c_{s_2}\sum_{t=1}^{T}(x_{jt}-w_{s_2t})^2-d_{s_2})^2 I(\chi_i=1,\xi_{is_2}=s_2)-\frac{1}{2\tau_6^2}c_{s_2}^2\right)I(c_{s_2}<0)$$

$$\propto \exp\left(-\frac{1}{2}\left\{\begin{array}{l}\left(\sum_{i=1}^{I}\sum_{s_2=1}^{S_2}\sum_{r=1}^{R}\sum_{j=1}^{J}\left(\sum_{t=1}^{T}(x_{jt}-w_{s_2t})^2\right)^2 I(\chi_i=1,\xi_{is_2}=s_2)+\frac{1}{\tau_6^2}\right)c_{s_2}^2 \\ -2\left(\sum_{i=1}^{I}\sum_{s_2=1}^{S_2}\sum_{r=1}^{R}\sum_{j=1}^{J}\left(\Delta_{ijr}-d_{s_2}\right)\left(\sum_{t=1}^{T}(x_{jt}-w_{s_2t})^2\right)I(\chi_i=1,\xi_{is_2}=s_2)\right)c_{s_2}\end{array}\right\}\right)I(c_{s_2}<0)$$

$$\sim N(\bar{c},V_c^{-1})I(c_{s_2}<0),$$

where:

$$V_c=\sum_{i=1}^{I}\sum_{s_2=1}^{S_2}\sum_{r=1}^{R}\sum_{j=1}^{J}\left(\sum_{t=1}^{T}(x_{jt}-w_{s_2t})^2\right)^2 I(\chi_i=1,\xi_{is_2}=s_2)+\frac{1}{\tau_6^2}$$

$$\bar{c}=\left(\sum_{i=1}^{I}\sum_{s_2=1}^{S_2}\sum_{r=1}^{R}\sum_{j=1}^{J}\left(\Delta_{ijr}-d_{s_2}\right)\left(\sum_{t=1}^{T}(x_{jt}-w_{s_2t})^2\right)I(\chi_i=1,\xi_{is_2}=s_2)\right)V_c^{-1}.$$

6. Generate additive parameter $d_{s_2}\mid\tau_3^2,\chi_i,\xi_{is_2},x_{jt},w_{s_2t},c_{s_2}$ for $s_2=1,\ldots,S_2$.

$$p(d_{s_2}\mid\sim)\propto\prod_{i=1}^{I}\left(\sum_{s_2=1}^{S_2}\left(\prod_{r=1}^{R}\prod_{j=1}^{J}f_{IP_{s_2}}(\Delta_{ijr}\mid x_{jt},w_{s_2t},c_{s_2})\right)I(\xi_{is_2}=s_2)\right)^{I(\chi_i=1)}p(d_{s_2})$$

$$\propto\prod_{i=1}^{I}\prod_{s_2=1}^{S_2}\prod_{r=1}^{R}\prod_{j=1}^{J}f_{IP_{S_2}}(\Delta_{ijr}\mid x_{jt},w_{s_2t},c_{s_2})^{I(\chi_i=1)I(\xi_{is_2}=s_2)}p(d_{s_2})$$

$$\propto\prod_{i=1}^{I}\prod_{s_2=1}^{S_2}\prod_{r=1}^{R}\prod_{j=1}^{J}\exp\left(-\frac{1}{2}(\Delta_{ijr}-c_{s_2}\sum_{t=1}^{T}(x_{jt}-w_{s_2t})^2-d_{s_2})^2\right)^{I(\chi_i=1)I(\xi_{is_2}=s_2)}\exp(-\frac{1}{2\tau_3^2}d_{s_2}^2)$$

$$\propto\exp\left(-\frac{1}{2}\sum_{i=1}^{I}\sum_{s_2=1}^{S_2}\sum_{r=1}^{R}\sum_{j=1}^{J}(\Delta_{ijr}-c_{s_2}\sum_{t=1}^{T}(x_{jt}-w_{s_2t})^2-d_{s_2})^2 I(\chi_i=1,\xi_{is_2}=s_2)-\frac{1}{2\tau_3^2}d_{s_2}^2\right)$$

$$\propto\exp\left(-\frac{1}{2}\left\{\begin{array}{l}\left[\sum_{i=1}^{I}\sum_{s_2=1}^{S_2}\sum_{r=1}^{R}\sum_{j=1}^{J}I(\chi_i=1,\xi_{is_2}=s_2)+\frac{1}{\tau_3^2}\right]d_{s_2}^2 \\ -2\left[\sum_{i=1}^{I}\sum_{s_2=1}^{S_2}\sum_{r=1}^{R}\sum_{j=1}^{J}\left(\Delta_{ijr}-c_{s_2}\sum_{t=1}^{T}(x_{jt}-w_{s_2t})^2\right)I(\chi_i=1,\xi_{is_2}=s_2)\right]d_{s_2}\end{array}\right\}\right)$$

$$\sim N(\bar{d},V_d^{-1}),$$

where:

$$V_d = \sum_{i=1}^{I}\sum_{s_2=1}^{S_2}\sum_{r=1}^{R}\sum_{j=1}^{J} I(\chi_i = 1, \xi_{is_2} = s_2) + \frac{1}{\tau_3^2}$$

$$\bar{d} = \left(\sum_{i=1}^{I}\sum_{s_2=1}^{S_2}\sum_{r=1}^{R}\sum_{j=1}^{J} \left(\Delta_{ijr} - c_{s_2} \sum_{t=1}^{T} (x_{jt} - w_{s_2 t})^2 \right) I(\chi_i = 1, \xi_{is_2} = s_2) \right) V_d^{-1}.$$

7. Generate the ideal point parameter $w_{s_2 t}$ for $s_2 = 1,\ldots,S_2$ and t = 1,…,T.

A random-walk Metropolis-Hastings algorithm is used to generate ideal point parameter $w_{s_2 t}$. Let $w_{s_2 t}^{(n)}$ denote a new candidate and $w_{s_2 t}^{(o)}$ represent the old value from the previous iteration of the chain. Draw a random vector (scalar) $w_{s_2 t}^{(n)} = w_{s_2 t}^{(o)} + \kappa e$, where κe is a draw from a candidate generating density N(0,κ).

Accept new vector $w_{s_2 t}^{(n)}$ with probability:

$$\alpha_w(w_{s_2 t}^{(n)}, w_{s_2 t}^{(o)}) = \min\left(\frac{p(w_{s_2 t}^{(n)})}{p(w_{s_2 t}^{(o)})}, 1 \right), \text{ and}$$

$$\frac{p(w_{s_2 t}^{(n)})}{p(w_{s_2 t}^{(o)})} \propto \frac{\prod_{i=1}^{I}\prod_{s_2=1}^{S_2}\prod_{r=1}^{R}\prod_{j=1}^{J} \left(f_{IP_{S_2}}(\Delta_{ijr} \mid x_{jt}, w_{s_2 t}^{(n)}, c_{s_2}, d_{s_2}) \right)^{I(\chi_i=1,\xi_{i_2}=s_2)} e^{-\frac{1}{2\tau_5^2}\left(w_{s_2 t}^{(n)}\right)^2} e^{-\frac{1}{2}\left(w_{s_2 t}^{(o)} - w_{s_2 t}^{(n)}\right)^2}}{\prod_{i=1}^{I}\prod_{s_2=1}^{S_2}\prod_{r=1}^{R}\prod_{j=1}^{J} \left(f_{IP_{S_2}}(\Delta_{ijr} \mid x_{jt}, w_{s_2 t}^{(o)}, c_{s_2}, d_{s_2}) \right)^{I(\chi_i=1,\xi_{i_2}=s_2)} e^{-\frac{1}{2\tau_5^2}\left(w_{s_2 t}^{(o)}\right)^2} e^{-\frac{1}{2}\left(w_{s_2 t}^{(n)} - w_{s_2 t}^{(o)}\right)^2}}$$

$$\propto \frac{\prod_{i=1}^{I}\prod_{s_2=1}^{S_2}\prod_{r=1}^{R}\prod_{j=1}^{J} \left(f_{IP_{S_2}}(\Delta_{ijr} \mid x_{jt}, w_{s_2 t}^{(n)}, c_{s_2}, d_{s_2}) \right)^{I(\chi_i=1,\xi_{i_2}=s_2)} e^{-\frac{1}{2\tau_5^2}\left(w_{s_2 t}^{(n)}\right)^2}}{\prod_{i=1}^{I}\prod_{s_2=1}^{S_2}\prod_{r=1}^{R}\prod_{j=1}^{J} \left(f_{IP_{S_2}}(\Delta_{ijr} \mid x_{jt}, w_{s_2 t}^{(o)}, c_{s_2}, d_{s_2}) \right)^{I(\chi_i=1,\xi_{i_2}=s_2)} e^{-\frac{1}{2\tau_5^2}\left(w_{s_2 t}^{(o)}\right)^2}}.$$

8. Generate the brand coordinate parameter x_{jt} for j = 1,…,J and t = 1,…,T.

As discussed in the identification section, brand coordinates can be separated into two parts: one with a constraint such that parameters need to be confined on the positive space and the other without this constraint. Let $x_{jt(nc)}$ be unconstrained brand coordinates and $x_{jt(c)}$ be constrained brand coordinates. For the proposed model, $x_{jt(nc)}$ and $x_{jt(c)}$ are randomly drawn from the respective posterior distribution iteratively and recursively as follows. First, a random-walk Metropolis Hastings algorithm with a normal prior $p(x_{jt(nc)}) \sim N(0, \tau_1^2)$ is used to generate unconstrained parameter $x_{jt(nc)}$. Let $x_{jt(nc)}^{(n)}$ denote a new candidate of the unconstrained parameter and $x_{jt(nc)}^{(o)}$ be previous draw of $x_{jt(nc)}$. A new candidate $x_{jt(nc)}^{(n)}$ is given by $x_{jt(nc)}^{(n)} = x_{jt(nc)}^{(o)} + \omega e$, where ωe is a draw from a candidate generating density $N(0, \omega)$. Here, ω needs to be calibrated to ensure more than 30% acceptance rate (Gelman, Gilks, and Roberts 1996). Accept the new candidate $x_{jt(nc)}^{(n)}$ with probability:

$$\alpha_{x_{nc}}(x_{jt(nc)}^{(n)}, x_{jt(nc)}^{(o)}) = \min\left(\frac{p(x_{jt(nc)}^{(n)})}{p(x_{jt(nc)}^{(o)})}, 1\right), \text{ and}$$

$$\frac{p(x_{jt(nc)}^{(n)})}{p(x_{jt(nc)}^{(o)})}=\left[\left(\frac{\prod_{i=1}^{I}\left[\left(\prod_{s_1=1}^{S_1}\prod_{r=1}^{R}\prod_{j=1}^{J} f_{v_{s_1}}(\Delta_{ijr}\mid x_{jt(nc)}^{(n)},v_{s_1t},b_{s_1})\right)^{I(\chi_i=0,\xi_{is_1}=s_1)}\times\left(\prod_{s_2=1}^{S_2}\prod_{r=1}^{R}\prod_{j=1}^{J} f_{IP_{s_2}}(\Delta_{ijr}\mid x_{jt(nc)}^{(n)},w_{s_2t},c_{s_2},d_{s_2})\right)^{I(\chi_i=1,\xi_{is_2}=s_2)}\right]}{\prod_{i=1}^{I}\left[\left(\prod_{s_1=1}^{S_1}\prod_{r=1}^{R}\prod_{j=1}^{J} f_{v_{s_1}}(\Delta_{ijr}\mid x_{jt(nc)}^{(o)},v_{s_1t},b_{s_1})\right)^{I(\chi_i=0,\xi_{is_1}=s_1)}\times\left(\prod_{s_2=1}^{S_2}\prod_{r=1}^{R}\prod_{j=1}^{J} f_{IP_{s_2}}(\Delta_{ijr}\mid x_{jt(nc)}^{(o)},w_{s_2t},c_{s_2},d_{s_2})\right)^{I(\chi_i=1,\xi_{is_2}=s_2)}\right]}\right)\times\left(\frac{e^{-\frac{1}{2\tau_1^2}\left(x_{jt(nc)}^{(n)}\right)^2}}{e^{-\frac{1}{2\tau_1^2}\left(x_{jt(nc)}^{(o)}\right)^2}}\right)\times\left(\frac{e^{-\frac{1}{2}\left(x_{jt(nc)}^{(o)}-x_{jt(nc)}^{(n)}\right)^2}}{e^{-\frac{1}{2}\left(x_{jt(nc)}^{(n)}-x_{jt(nc)}^{(o)}\right)^2}}\right)\right]$$

$$=\left(\frac{\prod_{i=1}^{I}\left[\left(\prod_{s_1=1}^{S_1}\prod_{r=1}^{R}\prod_{j=1}^{J} f_{v_{s_1}}(\Delta_{ijr}\mid x_{jt(nc)}^{(n)},v_{s_1t},b_{s_1})\right)^{I(\chi_i=0,\xi_{is_1}=s_1)}\times\left(\prod_{s_2=1}^{S_2}\prod_{r=1}^{R}\prod_{j=1}^{J} f_{IP_{s_2}}(\Delta_{ijr}\mid x_{jt(nc)}^{(n)},w_{s_2t},c_{s_2},d_{s_2})\right)^{I(\chi_i=1,\xi_{is_2}=s_2)}\right]}{\prod_{i=1}^{I}\left[\left(\prod_{s_1=1}^{S_1}\prod_{r=1}^{R}\prod_{j=1}^{J} f_{v_{s_1}}(\Delta_{ijr}\mid x_{jt(nc)}^{(o)},v_{s_1t},b_{s_1})\right)^{I(\chi_i=0,\xi_{is_1}=s_1)}\times\left(\prod_{s_2=1}^{S_2}\prod_{r=1}^{R}\prod_{j=1}^{J} f_{IP_{s_2}}(\Delta_{ijr}\mid x_{jt(nc)}^{(o)},w_{s_2t},c_{s_2},d_{s_2})\right)^{I(\chi_i=1,\xi_{is_2}=s_2)}\right]}\right)\times\left(\frac{e^{-\frac{1}{2\tau_1^2}\left(x_{jt(nc)}^{(n)}\right)^2}}{e^{-\frac{1}{2\tau_1^2}\left(x_{jt(nc)}^{(o)}\right)^2}}\right)$$

$$=\frac{L(\Delta_{ijr}\mid x_{jt(nc)}^{(n)},rest)}{L(\Delta_{ijr}\mid x_{jt(nc)}^{(o)},rest)}\times\left(\frac{e^{-\frac{1}{2\tau_1^2}\left(x_{jt(nc)}^{(n)}\right)^2}}{e^{-\frac{1}{2\tau_1^2}\left(x_{jt(nc)}^{(o)}\right)^2}}\right),$$

where *rest* means other parameters in the likelihood.

Next, a Metropolis Hastings algorithm with a Gamma prior $p(x_{jt(c)})\sim G(a,b)$ and a Gamma proposal is used to generate constrained parameters $x_{jt(c)}$. Let $x_{jt(c)}^{(n)}$ denote a new candidate of the constrained parameter and $x_{jt(c)}^{(o)}$ be previous draw of $x_{jt(c)}$. For the Gamma proposal, we reparameterize the shape parameter of the Gamma distribution kernel as $k\left(x_{jt(c)}^{(o)}\right)^2$ and the scale parameter as $\frac{1}{kx_{jt(c)}^{(o)}}$, so that the new candidate $x_{jt(c)}^{(n)}$ has the mean equal to the previous draw $x_{jt(c)}^{(o)}$ and the variance $1/k$ (Bradlow and Schmittlein 2000). Therefore, a new candidate $x_{jt(c)}^{(n)}$ is generated from $Gamma\left(k\left(x_{jt(c)}^{(o)}\right)^2,\frac{1}{kx_{jt(c)}^{(o)}}\right)$,

and k needs to be tuned to get an adequate acceptance rate. Accept the new candidate $x_{jt(c)}^{(n)}$ with probability:

$$\alpha_{x_c}(x_{jt(c)}^{(n)}, x_{jt(c)}^{(o)}) = \min\left(\frac{p(x_{jt(c)}^{(n)})}{p(x_{jt(c)}^{(o)})}, 1\right), \text{ and}$$

$$\frac{p(x_{jt(c)}^{(n)})}{p(x_{jt(c)}^{(o)})} = \left[\begin{array}{l} \left(\frac{\prod_{i=1}^{I}\left[\left(\prod_{s_1=1}^{S_1}\prod_{r=1}^{R}\prod_{j=1}^{J} f_{v_{s_1}}(\Delta_{ijr} \mid x_{jt(nc)}^{(n)}, v_{s_1 t}, b_{s_1})\right)^{I(\chi_i=0,\xi_{is_1}=s_1)} \times \left(\prod_{s_2=1}^{S_2}\prod_{r=1}^{R}\prod_{j=1}^{J} f_{IP_{s_2}}(\Delta_{ijr} \mid x_{jt(nc)}^{(n)}, w_{s_2 t}, c_{s_2}, d_{s_2})\right)^{I(\chi_i=1,\xi_{is_2}=s_2)}\right]}{\prod_{i=1}^{I}\left[\left(\prod_{s_1=1}^{S_1}\prod_{r=1}^{R}\prod_{j=1}^{J} f_{v_{s_1}}(\Delta_{ijr} \mid x_{jt(nc)}^{(o)}, v_{s_1 t}, b_{s_1})\right)^{I(\chi_i=0,\xi_{is_1}=s_1)} \times \left(\prod_{s_2=1}^{S_2}\prod_{r=1}^{R}\prod_{j=1}^{J} f_{IP_{s_2}}(\Delta_{ijr} \mid x_{jt(nc)}^{(o)}, w_{s_2 t}, c_{s_2}, d_{s_2})\right)^{I(\chi_i=1,\xi_{is_2}=s_2)}\right]}\right) \\ \times \left(\frac{\frac{1}{\Gamma(a)b^a}\left(x_{jt(c)}^{(n)}\right)^{a-1} e^{-\frac{x_{jt(c)}^{(n)}}{b}}}{\frac{1}{\Gamma(a)b^a}\left(x_{jt(c)}^{(o)}\right)^{a-1} e^{-\frac{x_{jt(c)}^{(o)}}{b}}}\right) \times \left(\frac{\frac{1}{\Gamma(a_2)b_2^{a_2}}\left(x_{jt(c)}^{(o)}\right)^{a_2-1} e^{-\frac{x_{jt(c)}^{(o)}}{b_2}}}{\frac{1}{\Gamma(a_1)b_1^{a_1}}\left(x_{jt(c)}^{(n)}\right)^{a_1-1} e^{-\frac{x_{jt(c)}^{(n)}}{b_1}}}\right) \end{array}\right]$$

$$= \frac{L(\Delta_{ijr} \mid x_{jt(c)}^{(n)}, rest)}{L(\Delta_{ijr} \mid x_{jt(c)}^{(o)}, rest)} \times \left(\frac{\frac{1}{\Gamma(a_2)b_2^{a_2}}\left(x_{jt(c)}^{(o)}\right)^{a_2-a} e^{-x_{jt(c)}^{(o)}(\frac{1}{b_2}-\frac{1}{b})}}{\frac{1}{\Gamma(a_1)b_1^{a_1}}\left(x_{jt(c)}^{(n)}\right)^{a_1-a} e^{-x_{jt(c)}^{(n)}(\frac{1}{b_1}-\frac{1}{b})}}\right),$$

where $a_1 = k\left(x_{jt(c)}^{(o)}\right)^2$, $b_1 = \frac{1}{kx_{jt(c)}^{(o)}}$, $a_2 = k\left(x_{jt(c)}^{(n)}\right)^2$, and $b_2 = \frac{1}{kx_{jt(c)}^{(n)}}$.

9. Update hyper parameter $\tau_1^2 \mid k_1, u_1, x_{jt}$ **.**

$$p(\tau_1^{-2} \mid -) \propto \frac{1}{\Gamma(k_1)u_1^{k_1}}\left(\tau_1^{-2}\right)^{k_1-1} e^{-1/\tau_1^2 u_1}\left(\tau_1^{-2}\right)^{\frac{1}{2}} e^{-1/2\tau_1^2 x_{jt}^2}$$

$$\propto \left(\tau_1^{-2}\right)^{(k_1+1/2)-1} \exp\left(-\left(\tau_1^{-2}\right)\Big/\left(\frac{1}{\frac{1}{u_1}+\frac{x_{jt}^2}{2}}\right)\right)$$

$$\sim Gamma(k_1+\frac{1}{2}, \frac{2u_1}{u_1 x_{jt}^2+2}).$$

Similarly,

$$p(\tau_2^{-2} \mid \sim) \sim Gamma(k_2+\frac{1}{2}, \frac{2u_2}{u_2 b_{s_1}^2+2})$$

$$p(\tau_3^{-2} \mid \sim) \sim Gamma(k_3+\frac{1}{2}, \frac{2u_3}{u_3 d_{s_2}^2+2})$$

$$p(\tau_4^{-2} \mid \sim) \sim Gamma(k_4+\frac{1}{2}, \frac{2u_4}{u_4 v_{s_1 t}^2+2})$$

$$p(\tau_5^{-2} \mid \sim) \sim Gamma(k_5+\frac{1}{2}, \frac{2u_5}{u_5 w_{s_2 t}^2+2})$$

$$p(\tau_6^{-2} \mid \sim) \sim Gamma(k_6+\frac{1}{2}, \frac{2u_6}{u_6 c_{s_2}^2+2}).$$

10. Update hyper parameter $\phi_0 \mid \chi_i, \alpha_0, \beta_o$ **.**

$$p(\phi_0 \mid \sim) \propto \prod_{i=1}^{I} \phi_0^{I(\chi_i=0)} (1-\phi_0)^{I(\chi_i=0)} \phi_0^{\alpha_0-1} (1-\phi_0)^{\beta_0-1}$$

$$\propto \phi_0^{\sum_{i=1}^{I} I(\chi_i=0)+\alpha_0-1} (1-\phi_0)^{\sum_{i=1}^{I} I(\chi_i=1)+\beta_0-1}$$

$$\sim Beta(\sum_{i=1}^{I} I(\chi_i = 0) + \alpha_0, \sum_{i=1}^{I} I(\chi_i = 1) + \beta_0).$$

11. Update hyper parameter $\gamma_{s_1} \mid \alpha_{s_1}, \xi_{is_1}$ **.**

$$p(\gamma_{s_1} \mid \sim) \sim Dirichlet(\alpha_{s_1} + \sum_{i=1}^{I} I(\xi_{is_1} = s_1)).$$

12. Update hyper parameter $\gamma_{s_2} \mid \alpha_{s_2}, \xi_{is_2}$ **.**

$$p(\gamma_{s_2} \mid \sim) \sim Dirichlet(\alpha_{s_2} + \sum_{i=1}^{I} I(\xi_{is_2} = s_2)).$$

References

Andrews, Rick L. and Ajay K. Manrai (1999), "MDS Maps for Product Attributes and Market Response: An Application to Scanner Panel Data," *Marketing Science*, 18 (4), 584-604.

Balabanov, Antoaneta and Andres M. Kanner (2004), "Unrecognized and Untreated: Preventing and Treating Depression in Patients With Epilepsy," *Psychiatric Times*, 23 (13), http://www.psychiatrictimes.com/article/showArticle.jhtml?articleId=175802494.

Berndt, Ernst R., Iain M. Cockburn, and Zvi Griliches (1996), "Pharmaceutical Innovations and Market Dynamics: Tracking Effects on Price Indexes for Antidepressant Drugs," *Brookings Papers on Economic Activity, Microeconomics*, 133-99.

Bettman, James R., Mary Frances Luce, and John W. Payne (1998), "Constructive Consumer Choice Processes," *Journal of Consumer Research*, 25 (3), 187-217.

Bőckenholt, I. and I. Bőckenholt (1991), "Constrained Latent Class Analysis: Simultaneous Classification and Scaling of Discrete Choice Data," *Psychometrika*, 56, 699-716.

Bőckenholt, I. and W. Gaul (1989), "Generalized Latent Class Analysis: A New Methodology for Market Structure Analysis," in Conceptual and Numerical Analysis of Data, O. Opitz, Ed. West Germany: *Springer-Verlag*.

Bradlow, Eric T. and David C. Schmittlein (2000), "The Little Engines That Could: Modeling the Performance of World Wide Web Search Engines," *Marketing Science*, 19 (1), 43-62.

Carroll, J. Douglas (1972), "Individual Differences and Multidimensional Scaling," in Multidimensional scaling; theory and applications in the behavioral sciences, Shepard and Roger N. and Romney and A. Kimball and Nerlove and Sara Beth, Eds. New York: *Seminar Press*.

---- (1980), "Models and methods for multidimensional analysis of preferential choice (or other dominance) data," in Similarity and Choice, E.D. Lantermann and H. Feger, Eds. Vienna: *Hans Huber Publishers*.

Chib, Siddhartha (2002), "Markov Chain Monte Carlo Methods," in Subjective and Objective Bayesian Statistics, S. James Press, Ed. 2 ed.: *Wiley*.

Chintagunta, Pradeep K. (1994), "Heterogeneous Logit Model Implications for Brand Positioning," *Journal of Marketing Research*, 31, 304-11.

Cowles, Mary Kathryn and Bradley P. Carlin (1996), "Markov Chain Monte Carlo Convergence Diagnostics: A Comparative Review," *Journal of the American Statistical Association*, 91 (434), 883-904.

DeSarbo, Wayne, Duncan K.H. Fong, John C. Liechty, and Jennifer Chang Coupland (2005), "Evolutionary Preferences/Utility Functions: A Dynamic Perspective," *Psychometrika*, 70 (1), 179-202.

DeSarbo, Wayne S. and J. Douglas Carroll (1985), "Three-Way Metric Unfolding via Alternating Weighted Least Squares," *Psychometrika*, 50 (3), 275-300.

DeSarbo, Wayne S., Daniel J. Howard, and Kamel Jedidi (1991), "MULTICLUS: A New Method For Simultaneously Performing Multidimensional Scaling And Cluster Analysis," *Psychometrika*, 56 (1), 121-36.

DeSarbo, Wayne S. and Kamel Jedidi (1995), "The Spatial Representation of Heterogeneous Consideration Sets," *Marketing Science*, 14 (3), 326-42.

DeSarbo, Wayne S., Youngchan Kim, and Duncan Fong (1999), "A Bayesian Multidimensional Scaling Procedure for the Spatial Analysis of Revealed Choice Data," *Journal of Econometrics*, 89, 79-108.

DeSarbo, Wayne S., Youngchan Kim, Michel Wedel, and Duncan K.H. Fong (1998), "A Bayesian Approach to the Spatial Representation of Market Structure from Consumer Choice Data," *European Journal of Operational Research*, 111, 285-305.

DeSarbo, Wayne S., Ajay K. Manrai, and Lalita A. Manrai (1994), "Latent Class Multidimensional Scaling: A Review of Recent Developments in the Marketing and Psychometric Literature," in Advanced Methods of Marketing Research, Richard P. Bagozzi, Ed. Cambridge: *Blackwell*.

DeSarbo, Wayne S., V. Ramaswamy, and Peter Lenk (1993), "A Latent Class Procedure for the Analysis of Two-way Compositional Data," *Journal of Classification*, 10, 159-94.

DeSarbo, Wayne S. and Vithala R. Rao (1986), "A Constrained Unfolding Methodology for Product Positioning," *Marketing Science*, 5 (1), 1-19.

---- (1984), "GENFOLD2: A Set of Models and Algorithms for the GENeral UnFOLDing Analysis of Preference/Dominance Data," *Journal of Classification*, 2, 147-68.

DeSarbo, Wayne S., Martin R. Young, and Arvind Rangaswamy (1997), "A Parametric Multidimensional Unfolding Procedure for Incomplete Nonmetric Preference/Choice Set Data in Marketing Research," *Journal of Marketing Research*, 34, 499-516.

Deun, K. Van, P.J.F. Groenen, W.J. Heiser, F.M.T.A. Busing, and L. Delbeke (2005), "Interpreting Degenerate Solutions In Unfolding By Use Of The Vector Model And The Compensatory Distance Model," *Psychometrika*, 70 (1), 45-69.

Diebolt, Jean and Christian P. Robert (1994), "Estimation of Finite Mixture Distributions through Bayesian Sampling," *Journal of the Royal Statistical Society. Series B (Methodological)*, 56 (2), 363-75.

Elrod, Terry (1988), "Choice Map: Inferring a Product Market Map From Observed Choice Behavior," *Marketing Science*, 7, 21-40.

Elrod, Terry and Michael P. Keene (1994), "A Factor-Analytic Model for Representing the Market Structure in Panel Data," *Journal of Marketing Research*, 32, 1-16.

Erdem, Tulin (1996), "A Dynamic Analysis of Market Structure Based on Panel Data," *Marketing Science*, 15 (4), 359-78.

Escobar, Michael D. and Mike West (1995), "Bayesian Density Estimation and Inference Using Mixtures," *Journal of the American Statistical Association*, 90 (430), 577-88.

Evans, Michael and Tim Swartz (1995), "Methods for Approximating Integrals in Statistics with Special Emphasis on Bayesian Integration Problems," *Statistical Science*, 10 (3), 254-72.

Gelfand, Alan E. and Adrian F. M. Smith (1990), "Sampling-based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association*, 85, 398-409.

Gelman, A., W.R. Gilks, and G.O. Roberts Eds. (1996), Efficient Metropolis Jumping Rules. Oxford, UK: *Oxford University Press*.

Gensch, Dennis H. (1987), "Empirical Evidence Supporting the Use of Multiple Choice Models in Analyzing a Population," *Journal of Marketing Research*, 24, 197-207.

Gilbride, Timothy J. and Greg M. Allenby (2004), "A Choice Model with Conjunctive, Disjunctive, and Compensatory Screening Rules," *Marketing Science*, 23 (3), 391-406.

Gilks, Walter R., Sylvia Richardson, and David J. Spiegelhalter (1996), "Introducing Markov Chain Monte Carlo," in Markov Chain Monte Carlo in Practice, W.R. Gilks and S. Richardson and D.J. Spiegelhalter, Eds.: *Chapman & Hall/CRC*.

Green, Peter J. (1995), "Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination," *Biometrika*, 82 (4), 711-32.

Hastings, W. K. (1970), "Monte Carlo Sampling Methods Using Markov Chains and Their Applications," *Biometrika*, 57 (1), 97-109.

Heiser, Willem J. (1989), "Order Invariant Unfolding Analysis Under Smoothness Restrictions," in Advances in Psychology, G.E. Stelmach and P.A. Vroon, Eds. Vol. 60: *North-Holland*.

Jedidi, Kamel and Rajeev Kohli (2005), "Probabilistic Subset-Conjunctive Models for Heterogeneous Consumers," *Journal of Marketing Research*, 42, 483-94.

Kamakura, Wagner A. (1991), "Estimating Flexible Distributions of Ideal-Points with External Analysis of Preferences," *Psychometrika*, 56 (3), 419-31.

Kamakura, Wagner A., Byung-Do Kim, and Jonathan Lee (1996), "Modeling Preference and Structural Heterogeneity in Consumer Choice," *Marketing Science*, 15 (2), 152-72.

Kass, Robert E. and Adrian E. Raftery (1995), "Bayes Factors," *Journal of the American Statistical Association*, 90 (430), 773-95.

Liechty, John C., Duncan K.H. Fong, and Wayne S. DeSarbo (2005), "Dynamic Models with Individual Level Heterogeneity: Applied to Evolution During Conjoint Studies," *Marketing Science*, 24 (2), 285-93.

Lopes, Hedibert Freitas (2000), "Bayesian Analysis in Latent Factor and Longitudinal Models," Duke University.

Manchanda, Puneet, Peter E. Rossi, and Pradeep K. Chintagunta (2004), "Response Modeling with Nonrandom Marketing-Mix Variables," *Journal of Marketing Research*, 41, 467-78.

Menza, Matthew (2006), "STAR*D: The Results Begin to Roll in," *American Journal of Psychiatry* 163, 1123.

Metropolis, M., A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller (1953), "Equations of state calculations by fast computing machine," *Journal of Chemical Physics*, 21, 1087-91.

Newton, Michael A. and Adrian E. Raftery (1994), "Approximate Bayesian Inference with the Weighted Likelihood Bootstrap," *Journal of the Royal Statistical Society. Series B (Methodological)*, 56 (1), 3-48.

Oh, Man-Suk and Adrian E. Raftery (2001), "Bayesian Multidimensional Scaling and Choice of Dimension," *Journal of the American Statistical Association*, 96 (455), 1031-44.

Psychiatric, Association American (2000), "Practice Guideline for the Treatment of Patients With Major Depressive Disorder."

Richardson, Sylvia and Peter J. Green (1997), "On Bayesian Analysis of Mixtures with an Unknown Number of Components," *Journal of the Royal Statistical Society. Series B (Methodological)*, 59 (4), 731-92.

Robert, Christian P. and George Casella (1999), Monte Carlo statistical methods. New York: *Springer*.

Roeder, Kathryn and Larry Wasserman (1997), "Practical Bayesian Density Estimation Using Mixtures of Normals " *Journal of the American Statistical Association*, 92 (439), 894-902

Rust, Roland T., Duncan Simester, Roderick J. Brodie, and V. Nilikant (1995), "Model Selection Criteria: An Investigation of Relative Accuracy, Posterior Probabilities, and Combinations of Criteria," *Management Science*, 41 (2), 322-33.

Sinha, Indrajit and Wayne S. DeSarbo (1998), "An Integrated Approach Toward the Spatial Modeling of Perceived Customer Value," *Journal of Marketing Research*, 35, 236-49.

Stephens, Matthew (2000), "Dealing with label-switching in mixture models," *Journal of the Royal Statistical Society, Series B*, 62, 795-809.

Sujan, Mita (1985), "Consumer Knowledge: Effects on Evaluation Strategies Mediating Consumer Judgments," *Journal of Consumer Research*, 12 (1), 31-46.

Tanner, Martin A. and Wing Hung Wong (1987), "The Calculation of Posterior Distributions by Data Augmentation," *Journal of the American Statistical Association*, 82 (398), 528-40.

Thase, Michael E., Edward S. Friedman, and Robert H. Howland (2000), "Venlafaxine and Treatment-Resistant Depression," *Depression and Anxiety*, 12 (Supplement 1), 55-62.

Wasserman, Larry. (2000), "Asymptotic inference for mixture models by using data-dependent priors," *Journal of the Royal Statistical Society: Series B (Methodological)*, 62 (1), 159-80.

Wedel, Michel and Wayne S. DeSarbo (1996), "An Exponential-Family Multidimensional Scaling Mixture Methodology," *Journal of Business & Economic Statistics*, 14 (4), 447-59.

Wedel, Michel and Wagner Kamakura (2000), Market Segmentation: Conceptual and Methodological Foundations: *Kluwer Academic Publishers*.

Winsberg, Suzanne and Geert De Soete (1993), "A Latent Class Approach to Fitting the Weighted Euclidean Model, CLASCAL," *Psychometrika*, 58 (2), 315-30.

Yang, Sha and Greg M. Allenby (2000), "A Model for Observation, Structural, and Household Heterogeneity in Panel Data," *Marketing Letters*, 11 (2), 137-49.

Young, Forrest W. (1987), Multidimensional Scaling: History, Theory, and Applications: *Lawrence Erlbaum Associates, Inc*.

JOONWOOK PARK

CURRICULUM VITAE

EDUCATION

Ph.D. Marketing (2007)
Smeal College of Business, Pennsylvania State University
M.S., Marketing Research (2000)
Business school, University of Wisconsin-Madison
M.B.A., Marketing (1996)
B.A., Business Administration (1993)
Business Administration, Korea University

HONERS AND AWARDS

- Ryder Fellowship, Pennsylvania State University, 2006
- Fellow at AMA Sheth Foundation Doctoral Consortium, 2006
- Fellow at Marketing Science Conference, 2006
- Anderson Scholarship, Pennsylvania State University, 2006
- Fellow at Haring Symposium, Indiana University, 2005
- Scott Award, Pennsylvania State University, 2005
- Finalist for the Marketing Science Practice Award, 2004
- McKinnon Graduate Award, University of Wisconsin-Madison, 1999
- A.C. Nielsen Scholarship, University of Wisconsin-Madison, 1999
- Graduate Scholarship, Korea University, 1994-1996

PUBLICATIONS

"A Choice Map approach to modeling Attribute-level Varied Behavior Among SKUs,"* Inman J. Jeffrey, Joonwook Park, and Ashish Sinha (Forthcoming), *Journal of Marketing Research*

* Equal Contribution

"Attribute Drivers: A Factor Analytic Choice Map Approach for Understanding Choices Among SKUs," Sinha Ashish, J. Jeffrey Inman, Yantao Wang, and Joonwook Park (2005), *Marketing Science*, Vol. 24, pp. 351-359

"Demand Forecasting with Early Stage Diffusion Data in Korean Cellular Market," Park Joonwook and Seoil Chaiy (1997), *Management Journal in Korea*, Vol. 41, pp. 199-222