A HIERARCHICAL BAYESIAN MULTIDIMENSIONAL SCALING METHODOLOGY FOR ACCOMMODATING BOTH STRUCTURAL AND PREFERENCE HETEROGENEITY

JOONWOOK PARK

SOUTHERN METHODIST UNIVERSITY

WAYNE S. DESARBO AND JOHN LIECHTY

PENNSYLVANIA STATE UNIVERSITY

Multidimensional scaling (MDS) models for the analysis of dominance data have been developed in the psychometric and classification literature to simultaneously capture subjects' *preference heterogeneity* and the underlying dimensional structure for a set of designated stimuli in a parsimonious manner. There are two major types of latent utility models for such MDS models that have been traditionally used to represent subjects' underlying utility functions: the scalar product or vector model and the ideal point or unfolding model. Although both models have been widely applied in various social science applications, implicit in the assumption of such MDS methods is that all subjects are homogeneous with respect to their underlying utility function; i.e., they all follow a vector model or an ideal point model. We extend these traditional approaches by presenting a Bayesian MDS model that combines both the vector model and the ideal point model in a generalized framework for modeling metric dominance data. This new Bayesian MDS methodology explicitly allows for mixtures of the vector and the ideal point models thereby accounting for both *preference heterogeneity* and *structural heterogeneity*. We use a marketing application regarding physicians' prescription behavior of antidepressant drugs to estimate and compare a variety of spatial models.

Key words: Bayesian multidimensional scaling, structural heterogeneity, preference heterogeneity, multidimensional unfolding model, multidimensional vector model, pharmaceutical marketing.

1. Introduction

Multidimensional scaling (MDS) refers broadly to a plethora of spatial models used to obtain multidimensional representations for the structure in various types of data including proximity, profile, and dominance data. Typically, the MDS analysis of dominance data is based on one of two distinct types of latent utility models: Slater's (1960) and Tucker's (1960) scalar product or vector model, and Coombs' (1964) ideal point or unfolding model. Although both models assume that subjects reach their preferences by considering a multidimensional set of stimulus characteristics, these models have rather different underlying utility assumptions (DeSarbo, Young, & Rangaswamy, 1997). In the simple ideal point or unfolding model, utility decreases for a subject for stimulus locations farther away from a subject's ideal point in any direction. For the two-dimensional ideal point utility model, the isopreference contours are circles of equal radii around the subjects' ideal points. While the ideal point model is a more generalized case of the vector model (Carroll, 1972), it often suffers from degenerate solutions that hinders one from interpreting and using the derived solutions (see Busing, Groenen, & Heiser (2005) for a

recent discussion of this issue in ordinary multidimensional unfolding analysis as well as a literature review of procedures devised to remedy degenerate solutions). The vector MDS model represents subjects by vectors and stimuli by coordinates. Here, subjects' preferences are modeled by the orthogonal projection of the stimulus coordinates onto these subject vectors. As such, higher values of projection indicate higher utility. The underlying assumption of the vector utility model is "the more the better": stimuli positioned farther out in the direction of a subject's vector have higher predicted utility. Therefore, the isopreference contours in two dimensions are straight lines perpendicular to the subject's vector.

Although these two types of spatial models have been separately applied in various social science applications, little attention has been given to comparative *structural heterogeneity* in contrast to the vast literature on *preference heterogeneity*, where individuals have different preference parameters conditional on a specific utility model. By *structural heterogeneity*, we refer to differences in the structure of underlying decision processes (Kamakura, Kim, & Lee, 1996). As such, implicit in the assumptions of existing MDS methods are that all subjects are homogeneous with respect to the underlying utility function used in their preference formulation (either all unfolding or all scalar products). Recent research, however, suggests that this may not be the case. Deun, Groenen, Heiser, Busing, and Delbeke (2005) contended, in the context of traditional MDS, 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 stimuli coordinates can be replaced by vectors without altering subjects' preference order. This implies that combinations of the vector model and the ideal point model need to be explicitly considered, and that a generalized approach which can explicitly identify "mixtures of vector and unfolding (ideal point) representations" (DeSarbo & Carroll, 1985) would be an important contribution to the MDS literature. Furthermore, several studies have shown that a subject can exhibit changes in decision over time, situation, occasion, or context (Belk, 1975; Petty & Cacioppo, 1986; Srivastava, Alpert, & Shocker, 1984). Both the psychology and consumer behavior literature support the notion that individuals do *not* typically have stable utility functions. Instead, they construct their final utility assessments spontaneously as they face specific decision problems. Moreover, these on-the-spot judgments are shaped by the needs and goals of the individual, the aspects of the situation, the context in which the choice alternatives are being evaluated, how the choice problem is stated, and how the choice alternatives are framed (Bettman, Luce, & Payne, 1998; Belk, 1974, 1975, 1979). Thus, subsequent decisions are a function of *both* the context in which the decisions are made and the individual making the decision (see Tversky & Kahneman, 1991; Simon, 1955, 1990). Hence, individuals can employ a different decision strategy each time they make a decision. As a result, individuals may have different preference judgments for the same stimuli across different situations or contexts (Bettman et al., 1998), and context or situational effects may affect subjects' decision making in different situations (see Belk, 1975; Srivastava et al., 1984).

This manuscript presents a generalized MDS model that explicitly accommodates *both* *structural* and *preference* heterogeneity. Specifically, we propose a Bayesian MDS model that 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-only and the ideal point-only model can be considered as special cases of the proposed model. Second, within each structure (type of utility function representation), *preference heterogeneity* is explicitly accommodated since estimation is performed at the individual level. Third, we extend structural heterogeneity to the situation level so that a subject may exhibit mixtures of the vector and the ideal point depending upon the situation or context by relaxing the assumption that all observations from a subject are homogeneous. Finally, inference is performed using Markov chain Monte Carlo (MCMC)

methods, which are logically developed given the hierarchical model specification that reflects the uncertainty in the model parameters.

We proceed by first defining the vector and ideal point MDS model which we then generalize to the proposed model. We 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, likelihood function, joint posterior distribution, and an estimation procedure using a MCMC algorithm to accommodate *preference heterogeneity*. Detailed descriptions of the full conditional distributions are also provided. In Section 3, we present a marketing application involving an analysis of physician's prescribing behavior of several brands of antidepressants over time. Here, we demonstrate the ability of the proposed methodology to portray individual physicians who "switch" between vector and ideal point latent utility functions by time period. We conclude the paper with a discussion of future research directions.

2. The Proposed Model

For $i = 1, \ldots, I$ subjects who make preference judgments toward $j = 1, \ldots, J$ stimuli in $r = 1, \ldots, R$ situations (e.g., time periods, experimental treatments, occasions, replications, etc.), let Δ_{ijr} denote the preference rating (or dominance score) for stimulus j by subject i in the r-th situation; and $t = 1, \ldots, T$ unknown dimensions. We define the corresponding latent vector and the ideal point utility functions as:

$$U_{ijr|V} = \sum_{t=1}^{T} x_{jt} v_{it} + b_i + e_{ijr|V}, \tag{1}$$

$$U_{ijr|IP} = c_i \sum_{t=1}^{T} (x_{jr} - w_{it})^2 + d_i + e_{ijr|IP}, \tag{2}$$

where $e_{ijr|V} \sim N(0, \delta_V^2)$ and $e_{ijr|IP} \sim N(0, \delta_{IP}^2)$. Here, $U_{ijr|V}$ and $U_{ijr|IP}$ represent subject i's latent utility toward stimulus j in situation r given that the subject belongs to a vector model or the ideal point model, respectively. Without loss of generality, we assume that $\delta_V^2 = \delta_{IP}^2 = 1$. In (1) and (2), x_{jt} represents stimulus j's coordinate on the t-th dimension, and b_i and d_i represent additive constants for the vector model and the ideal point model, respectively. v_{it} represents the t-th coordinate of subject i's vector, and w_{it} represents the t-th coordinate of the ideal point for individual i. Finally, c_i is a scale parameter affecting the squared Euclidean distance between stimulus coordinates and ideal points of subject i. As such, (1) and (2) represent typical vector and ideal point utility functions, respectively. As indicated in GENFOLD2 (DeSarbo & Rao, 1984, 1986), a positive sign for scale parameter c_i indicates an anti-ideal point when the data represent preference. For ease of interpretation, this scale parameter c_i is constrained to be strictly negative so as to represent an ideal point model. We impose a restriction for this scale parameter in the section of the paper describing the prior distributions.

Given the utility specified in (1) and (2), the individual likelihood functions for the vector and the ideal point model are constructed as follows:

$$L_{i|V} = \prod_{r=1}^{R} \prod_{j=1}^{J} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\Delta_{ijr} - \sum_{t=1}^{T} x_{jt} v_{it} - b_i\right)^2\right), \tag{3}$$

$$L_{i|IP} = \prod_{r=1}^{R} \prod_{j=1}^{J} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\Delta_{ijr} - c_i \sum_{t=1}^{T}(x_{jt} - w_{it})^2 - d_i\right)^2\right), \tag{4}$$

where $L_{i|V}$ and $L_{i|IP}$ represent individual level likelihoods given that the observations from a subject i belong to the vector model or the ideal point model respectively.

Note, (3) and (4) are not much different from existing MDS models that accommodate *preference heterogeneity* via vectors v_{it} and ideal points w_{it}. We now incorporate *structural heterogeneity* by introducing a latent variable χ_i such that $P(\chi_i = V) = \phi_1$ and $P(\chi_i = IP) = 1 - \phi_1$. This latent variable determines the probability that subject i belongs to either a vector model (i.e., ϕ_1) or an ideal point model (i.e., $1 - \phi_1$). The resulting individual level *complete likelihood* is:

$$L_i = L_{i|V}\phi_1 + L_{i|IP}(1 - \phi_1). \tag{5}$$

Following the standard data augmentation procedure (Diebolt & Robert, 1994; Tanner & Wong, 1987), we augment this structural heterogeneity latent variable ϕ_1 and $1 - \phi_1$ using an indicator function such as $I(\chi_i = V)$ and $I(\chi_i = IP) = 1 - I(\chi_i = V)$. Therefore (5) can be rewritten as

$$L_i = L_{i|V}I(\chi_i = V) + L_{i|IP}I(\chi_i = IP). \tag{6}$$

Note that the specification of (6), which we will refer to as *Model 1*, is a generalized model that nests both the vector and the ideal point-only models. If no subject is assigned to the vector model, then this model becomes the ideal point model, and vice versa. Otherwise, a subset of subjects can be classified to the vector model, and the others to the ideal point model. Although Model 1 incorporates *structural heterogeneity* at the individual level, one can argue that this specification is somewhat restrictive as it does not permit a subject to exhibit heterogeneous utility in different situations. For instance, a subject can exhibit vector utility in some situations and ideal point utility in other situations. To incorporate this situation level structural heterogeneity, we extend Model 1 by introducing a situation level structural heterogeneity latent variable such that $P(\chi_{ir} = V) = \phi_2$ and $P(\chi_{ir} = IP) = 1 - \phi_2$. Similar to (6), the resulting individual level complete likelihood now becomes

$$L_i = \prod_{r=1}^{R}\left\{L_{ir|V}I(\chi_{ir} = V) + L_{ir|IP}I(\chi_{ir} = IP)\right\}, \tag{7}$$

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

$$L_{ir|V} = \prod_{j=1}^{J} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\Delta_{ijr} - \sum_{t=1}^{T} x_{jt}v_{it} - b_i\right)^2\right), \tag{8}$$

$$L_{ir|IP} = \prod_{j=1}^{J} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\Delta_{ijr} - c_i \sum_{t=1}^{T}(x_{jt} - w_{it})^2 - d_i\right)^2\right). \tag{9}$$

We will label this model extension as *Model 2*. Note that we do not have a situation-level vector or ideal point in (7) as situation-level parameters would require extensive data replications (situations). Rather, we try to relax Model 1's assumption that all observations of an individual are homogeneous. As such, we expect that the specification in (7) could capture structural changes in subjects' observations. Thus far, we have postulated two MDS models with both *preference heterogeneity* and *structural heterogeneity* at the individual and occasion level. Note that the

specification of (7), i.e., Model 2, is a generalized version of Model 1; Model 2 is equivalent to Model 1 if all observations of a subject are classified to either the vector or the ideal point model.

To estimate the proposed models, we employ a hierarchical Bayesian approach. We use Markov chain Monte Carlo methods (MCMC) to generate random deviates from the posterior distributions without requiring analytic integration (Chib, 2002; Gelfand & Smith, 1990; Gilks, Richardson, & Spiegelhalter, 1996; Hastings, 1970; Metropolis, Rosenbluth, Rosenbluth, Teller, & Teller, 1953; Tanner & Wong, 1987). Next, we turn the discussion to the specification of the prior distributions and issues related to the identification of these MDS models.

2.1. Selection of the Prior Distributions

First, we assume a priori, that the structural heterogeneity parameter ϕ_k follows a Beta prior with parameters a_k and b_k for $k = 1, 2$, where $f(\phi_k) = \frac{1}{B(a_k, b_k)} \phi_k^{a_k-1} (1 - \phi_k)^{b_k-1}$ if $0 < \phi_k < 1$ and 0 otherwise. Here we set $a_k = b_k = 1$ to reflect the case of a lack of information. With this prior setting, the Beta prior becomes a Uniform distribution so that each individual (or occasion) has an equal chance to belong to either the vector or the ideal point model. We then assign a univariate Normal distribution with zero mean as a prior for all other parameters. That is, $P(x_{jt}) \sim N(0, \tau_x^2)$, $P(w_{it}) \sim N(0, \tau_w^2)$, $P(v_{it}) \sim N(0, \tau_v^2)$, $P(b_i) \sim (0, \tau_b^2)$, $P(d_i) \sim N(0, \tau_d^2)$, and $P(c_i) \sim N(0, \tau_c^2) I(c_i < 0)$, respectively.[1] As discussed before, the scale parameter c_i 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. We use the conjugate prior for the inverse of the variance $P(\tau_x^{-2}) \sim G(k_x, u_x)$, $(\tau_v^{-2}) \sim G(k_v, u_v)$, $P(\tau_w^{-2}) \sim G(k_w, u_w)$, $P(\tau_b^{-2}) \sim G(k_b, u_b)$, $P(\tau_d^{-2}) \sim G(k_d, u_d)$, and $P(\tau_c^{-2}) \sim G(k_c, u_c)$, respectively, where G denotes the Gamma distribution, and we set $k_x = \cdots = k_c = u_x = \cdots = u_c = 0.5$ to reflect the lack of prior information.

2.2. Issues in Identification

While specifying these models, we have acted as if both the vector and ideal point models are identifiable. However, this is not the case here as these spatial models are "under-identified"— these models have an infinite number of solutions that render the same likelihood values. Thus, parameters need to be constrained to obtain unique solutions (Wedel & DeSarbo, 1996; Wedel & Kamakura, 2000; Young, 1987). These "parameter indeterminacies" stem from the fact that orthogonal transformations of respective configurations in both vector and unfolding model portions do not alter the likelihood, seriously affecting the resulting inference from the proposed MCMC procedure. In addition, such spatial models can be translated, expanded, or reflected (depending upon either the vector or unfolding portion) without any effect on the likelihood (DeSarbo, Manrai, & Manrai, 1994; DeSarbo & Rao, 1984, 1986; Wedel & DeSarbo, 1996; Young, 1987). Without explicit constraints, summarizing the MCMC output will be hopelessly and unnecessarily complicated. Generally, these indeterminacies are acknowledged by subtracting the corresponding degrees of freedom from the number of model parameters in a maximum likelihood approach (DeSarbo & Cho, 1989). In Bayesian analysis, however, these indeterminacy issues have to be circumvented either by the imposition of strong or informative priors on stimulus coordinates (DeSarbo, Kim, & Fong, 1998, DeSarbo, Kim, Wedel, & Fong, 1999) or by a post-processing of the resulting coordinates (Oh & Raftery, 2001). A notable exception is the Bradlow & Schmittlein (2000) Mahalonobis distance formulation. To remove the rotational and reflection indeterminacies in their formulation, they imposed $2T$ constraints of which T stimulus coordinates are fixed at zero, and T other coordinates are constrained to lie in the positive

---

[1]We also tested shrinkage effects for additive parameters b_i and d_i such as $P(b_i) \sim N(\bar{b}, \tau_b^2)$ and $P(d_i) \sim N(\bar{d}, \tau_d^2)$, and found no significant difference in the results.

orthant of the derived space. In a similar vein, 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 equal to one for the first T stimulus coordinates, and constrain the remaining T parameters of these stimulus coordinates on the positive real line for our mixture specification of vector and ideal point representations. For instance, let $\underline{x} = (x_{11}, x_{21}, \ldots, x_{jt})' = (\underline{x}_c, \underline{x}_{nc})$, where $\underline{x}_c$ are constrained and $\underline{x}_{nc}$ are unconstrained stimulus 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 in two-dimensional space, and a similar approach can be taken in higher space. With these constraints, the reflection, rotational, and scale indeterminacies can be treated simultaneously. Finally, for the origin indeterminacy present in the unfolding model, we impose a restriction on the stimulus locations x_{jt} such that $\sum_{j=1}^{J} x_{jt} = 0$ for each dimension t ($t = 1, \ldots, T$).

2.3. Estimation Procedure and Model Selection

Given the likelihood and prior specification discussed earlier, we employ Markov chain Monte Carlo (MCMC) methods to iteratively generate samples from the posterior density (Gilks et al., 1996). Random starting values are generated for initial parameter values followed by the steps described in the Appendix. This process is iterated first for a designated number of burn-in cycles, and then for a preset number of estimation iterations.

For model and dimensionality selection, we select the best model by computing the posterior densities of the different specifications for the observed data. An alternative, which we do not implement but leave as a topic for future research, is to treat the dimensionality (t) as a random variable, and then to find posterior probabilities for t according to the proportion of time that the MCMC algorithm spends in dimension t (Green, 1995; Richardson & Green, 1997). While the second approach has its own merits, it is computationally very challenging, particularly with regards to the mixing properties of the resulting MCMC Markov chain. As such, we choose to estimate each model separately given fixed t, and compare the posterior densities of these specifications. For model selection criteria, we initially use Harmonic mean, $\hat{p}_1(D)$, and Newton and Raftery's fourth estimate, $\hat{p}_4(D)$ (Kass & Raftery, 1995; Newton & Raftery, 1994).

Although $\hat{p}_1(D)$ and $\hat{p}_4(D)$ are easy to implement, they often tend to favor 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. As such, we also use two additional criteria for model selection. First, we employ Bayesian Information Criteria (BIC) as it can be used as an approximation of the marginal likelihood. Rust, Simester, Brodie, and Nilakant (1995) compared various model selection heuristics and concluded that the BIC is the most consistent and accurate model selection criteria. We also use a model selection criterion similar to the reversible jump MCMC methods (referred to RJMCMC hereafter). 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 conducting RJMCMC in our context is to use a "mini" RJMCMC similar to a strategy described by Lopes (2000). The basic premise of this approach is to use the posterior likelihood at each iteration for each model $k = 1, \ldots, K$. 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 $P(k|k') = J(k \to k')$. 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., $p(k'|k) = \frac{I(k' \neq k)}{K-1}$ and $\frac{P(k'|k)}{P(k|k')} = 1$).

Step 2. Propose a new candidate Θ_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 $j \neq k, k'$ where D denotes data.

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

$$\alpha = \text{Min}\left(1, \left(\frac{P(k|k')P(D|\Theta_k)P(\Theta_k)P(D|\Theta_{k'})P(\Theta_{k'})P(D|\Theta_k^n)P(\Theta_{k'}^n)P(\Theta_k^n)\prod_{j\neq k,k'}P(\Theta_j)P(k)}{P(k|k')P(D|\Theta_k^n)P(\Theta_k^n)P(D|\Theta_{k'}^n)P(\Theta_{k'}^n)(D|\Theta_k)P(\Theta_k)P(\Theta_{k'})\prod_{j\neq k,k'}P(\Theta_j)P(k')}\right)\right)$$

$$= \text{Min}\left(1, \frac{P(D|\Theta_{k'})P(k')}{P(D|\Theta_k)P(k)}\right).$$

Thus, to determine which model provides the best fit to the data, we compute and inspect $\hat{p}_1(D)$, $\hat{p}_4(D)$, BIC, and RJMCMC model selection heuristics. Note, we generated synthetic data sets with known structural and preference heterogeneity to verify the proposed model's ability to uncover the true preference structure and correctly detect both types of heterogeneity, as well as to understand whether these various model selection criteria identify the true (known) structure. Results indicate that the proposed model does recover the true structure with high accuracy, and that all model selection heuristics point to the true model structure with respect to the synthetic data tested. Details of the simulation study are available online in a supplementary appendix.

3. A Pharmaceutical Marketing Application

A major US pharmaceutical company conducted a market research study among physicians (i.e., general practitioners, internists, psychiatrists, etc.) in order to understand their prescribing decisions for prescription antidepressant medications. Antidepressants are prescribed for symptoms such as depression, social anxiety disorder, and generalized anxiety disorder (GAD). 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 six months. Symptoms of depression often 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 (Karasu, Gelenberg, Merriam, & Wang, 2006). To be diagnosed as having major depression, a person must show at least five of the above symptoms. 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). Currently, the depression therapy market is one of the largest medication markets in the world. It is estimated that approximately 6% of US population—some 19 million people—will have a depressive illness that warrants treatment (Consumer Reports, 2005). In 2004, global sales of branded antidepressants exceeded $14 billion, and US sales totaled $9.9 billion.

The data are composed of prescriptions from a sample of 250 US physicians for five leading brands of antidepressants. These data concern the total number of prescriptions written for each brand recorded on a monthly basis over a seven-month period. Due to confidentiality agreements with the client pharmaceutical company, the specific brands are disguised and labeled with letters A to E. The leading five brands in the data can be categorized into three types of antidepressants based on their chemical components: (1) SSRI (selective serotonin reuptake inhibitors); (2) SNRI (serotonin and norepinephrine reuptake inhibitors); and (3) NDRI (norepinephrine and dopamine reuptake inhibitors). 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, including Brand B, Brand C, and Brand D, belong to this category. SNRIs are believed to work especially well for patients (up to 40%) 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. Finally, NDRI is a selective catecholamine (norepinephrine and dopamine)

reuptake inhibitor, and it has only a minor effect on serotonin reuptake. Brand E is an NDRI. These five brands can also be classified by their intended treatment.

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. In addition, many antidepressants often interact with other medications and caution must be exercised in their prescription to patients.

Brand-specific characteristics merit further explanation. First, Brand B was introduced more recently compared to other brands. Brand B also has a smaller chance of unwanted 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.[2] Compared to other brands, Brand D is the only brand approved for obsessive-compulsive disorder (OCD) in children and adolescents age 6–17 years. 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 Reports (2005) reported that Brand E has the lowest sexual dysfunction side effects based on 1664 patients' clinical results. However, if this medication is taken in increased dosage (e.g., 450 mg/day), Brand E has a higher risks of seizure. Furthermore, Brand E is associated with the development of some psychotic symptoms, including delusions and hallucinations, and is recommended for 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[3] (5 hours vs. 24 hours of other antidepressants). Note, antidepressants need to be taken for at least four to eight weeks before the treatment can be assessed. 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, & Griliches, 1996). The American Psychiatric Association's medical practice guidelines for the prescription of antidepressants include: (1) anticipated side effects and their safety or tolerability; (2) history of prior response of patient or family member; (3) patient preference; (4) cost; and (5) quantity and quality of clinical trial data. Doctors usually recommend an antidepressant that is least likely to cause side effects for the person taking it. A recent study led by the 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 (Karasu et al., 2006). It should be noted that there is no further information provided in these data whether each prescription represents a new prescription, renewal, or mixing prescriptions of multiple brands. Nor is there any individual difference information collected on these various physicians relating to type of physician, size of the practice, hospital-based vs. private practice, years of experience, geographic location, gender, etc.

The data are composed of a total of 162,515 prescriptions over a seven-month period. The prescription shares are: Brand A (22.2%), Brand B (21.3%), Brand C (16.3%), Brand D (16.4%),

---

[2]Brand A and Brand D are also known to have few drug interactions (see the Antidepressant Comparison Chart at www.RxFiles.ca).

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

TABLE 1.
Descriptive statistics on the total number of prescriptions.

	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

TABLE 2.
Comparison of model selection criteria.

Dimension		Calibration data						Hold out data	
	Model	Newton and raflery	Harmonic mean	BIC	RJMCMC	RMSE	VAF	RMSE	VAF
1	Vector only	−221398	−221365	444685	0	7.563	0.393	8.017	0.325
	Ideal point only	−259068	−258369	518693	0	8.189	0.295	8.470	0.246
	Model 1	−218123	−218017	439927	0	7.503	0.402	8.002	0.327
	Model 2	−207135	−206560	417012	0	7.297	0.439	7.908	0.345
2	Vector only	−135021	−134970	272878	0	5.844	0.638	6.859	0.510
	Ideal point only	−134836	−134785	272503	0	5.840	0.639	6.904	0.505
	Model 1	−122957	−122909	251662	0	5.562	0.672	6.854	0.514
	Model 2	**-105069**	**-105000**	**215846**	**1**	**5.115**	**0.725**	6.765	0.527
3	Vector only	−132778	−132715	269354	0	5.725	0.655	6.858	0.512
	Ideal point only	−132641	−132589	269100	0	5.722	0.655	6.898	0.509
	Model 1	−123385	−123085	253169	0	5.557	0.681	**6.648**	**0.536**
	Model 2	−106839	−106139	220078	0	5.144	0.724	6.694	0.534

and Brand E (23.9%), respectively. Table 1 shows the 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 a seven-month period (although there is substantial variation among this sample of physicians).

We initially preprocessed these prescription data by double mean centering to minimize the predominance of rows (i.e., prescription volume) and column (i.e., brand share) effects in the resulting spatial solution (see Harshman & Lundy, 1984a, 1985). We then calibrated the proposed model based on the first six months' prescriptions (140,404 prescriptions), and used the last month's prescriptions for model validation (22,111 prescriptions). Note that we set $c_i = -1$ to facilitate the interpretation of the derived joint space with respect to estimating ideal points (i.e., no anti-ideal points). To find the best-fitting model, we first estimated models without structural heterogeneity (i.e., the vector-only and ideal point-only models). Next, we compared them with the individual level structural heterogeneity model (Model 1), and finally the situation level structural heterogeneity model (Model 2). Table 2 presents results of the various model-selection criteria. The results in Table 2 suggest that the two-dimensional solution is most parsimonious for these data across the various model-selection criteria for models with structural heterogeneity. In particular, the two-dimensional solution for Model 2, which incorporates situation-level structural heterogeneity, shows the highest harmonic mean (−105,000) and Newton and Raftery's estimator (−105,069), as well as minimum BIC (215,846).
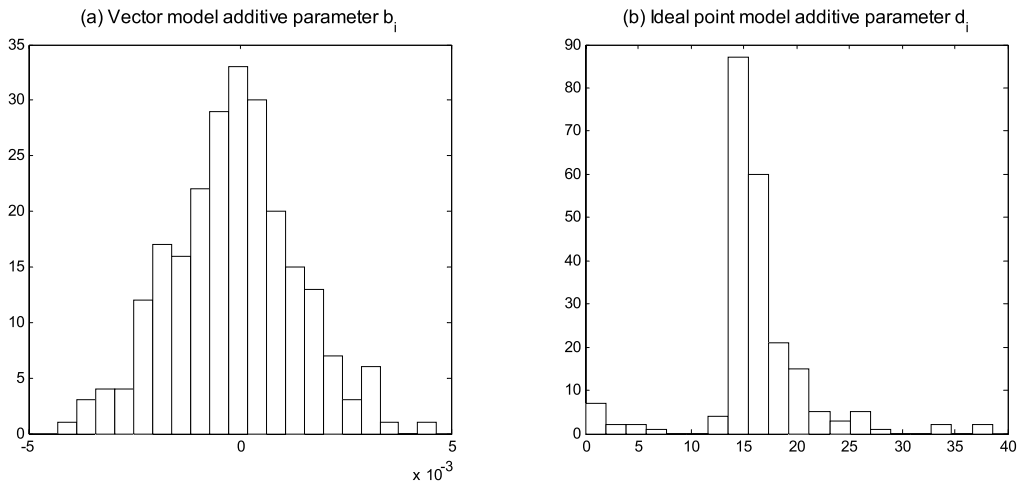
FIGURE 1.
Distribution of additive parameters.

A simple comparison of the results in Table 2 reveals that models that combine *both* structural and preference heterogeneity clearly outperform models only with preference heterogeneity regarding *all* model-selection criteria. In addition, other model solutions appear to require a third dimension. This is surprising as models without structural heterogeneity have more parameters than models with structural heterogeneity. For instance, the three-dimensional vector model has about 1.5 times as many parameters as Model 1 due to the increase in the dimensionality. It appears that the incorporation of structural heterogeneity results in the reduction of dimensionality. (We also experienced this same phenomena with our simulations of synthetic data whose structure is known. See the Web Appendix.) In addition to these model-selection criteria, we compared the model-selection probability using the "mini" RJMCMC method discussed earlier. Here, we ran 10 million iterations with likelihoods of the models estimated in order to get marginal posterior probability from the "mini" RJMCMC method, and used the last 5 million iterations to calculate the model-selection probability. As shown in Table 2, the "mini" RJMCMC method also points to the two-dimensional, occasion level structural heterogeneity model with the highest probability to be chosen (Prob.=1). Similar results can be found in such measures as Variance Accounted For (VAF) and Root Mean Squared Error (RMSE) for the calibration sample as shown in Table 2. Interestingly, for the validation hold-out sample, the three-dimensional structural heterogeneity models slightly outperform the two-dimensional structural heterogeneity solution. We will focus our discussion on the model with the most explanatory power as this allows us to facilitate the discussion with a two-dimensional joint space.

Next, we devote the remainder of this section to discussing the results of the two-dimensional, situation-level structural heterogeneity MDS solution. Here, we start our discussion with parameters of somewhat less interest. Figure 1 displays the distribution of additive parameters (e.g., b_i, d_i), and Table 3 shows summary statistics of hyperparameters. Note that we specify a hyper prior distribution for each of τ^2. As shown in Table 3, the differences in the posterior estimates for these hyperparameters vary dramatically, indicating that this additional level of model hierarchy can accommodate different levels of uncertainty. For instance, τ_b^2 and τ_d^2 exhibit huge difference in posterior means. We witness a similar result for two additional parameters, b_i and d_i, as shown in Figure 1. While the vector model additive parameter b_i is mostly concentrated around zero (mean $= -0.0002$, standard deviation $= 0.0002$), the ideal point additive parameter shows higher uncertainty (mean $= 16.04$, standard deviation $= 5.08$).

TABLE 3.
Summary of posterior estimates of hyperparameters.

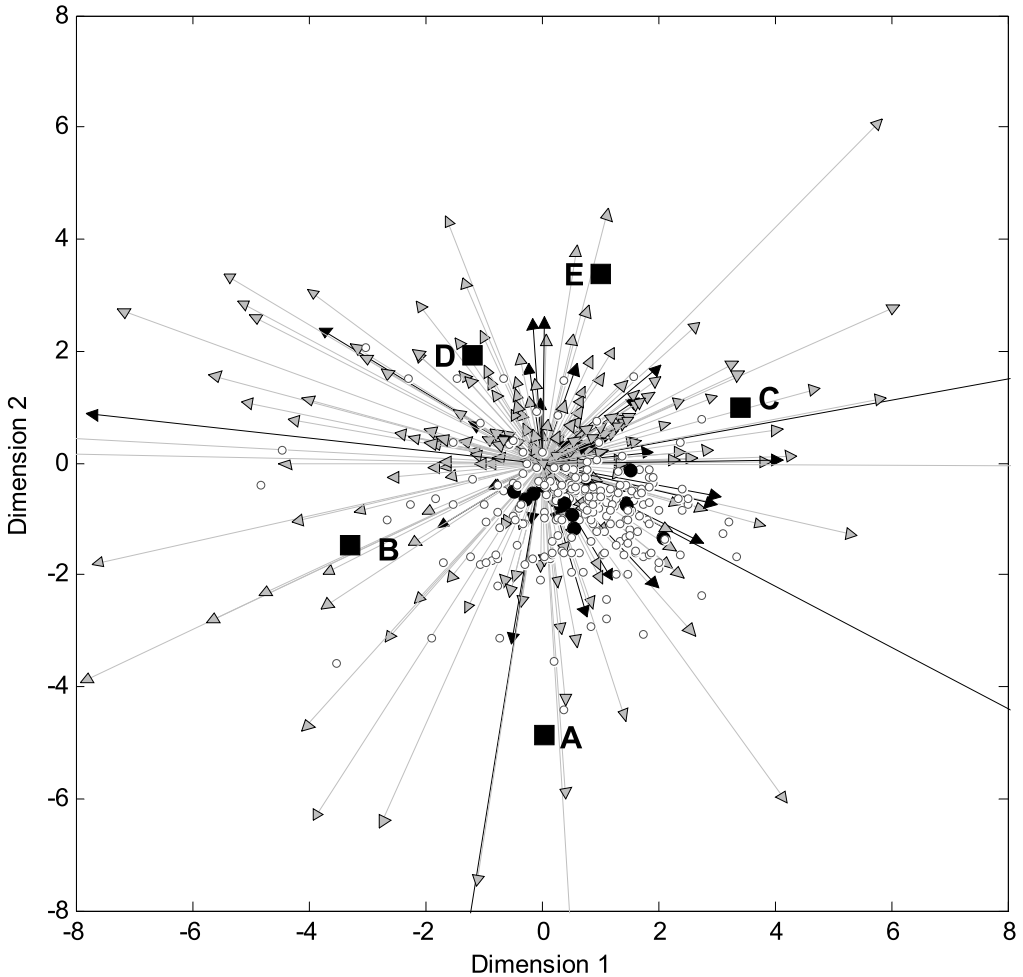|	τ_x^2	τ_b^2	τ_d^2	τ_v^2	τ_w^2	ϕ_2
Mean	9.20	0.04	300.29	5.77	1.73	0.62
Standard deviation	8.19	0.00	30.06	0.37	0.13	0.03

FIGURE 2.
Derived joint space for the two-dimensional Model 2. *Note*: individuals with only vector utility are represented as *solid black* vectors, and individuals with only ideal point utility are represented as *black circles* while individuals with both vector and ideal point utility are represented as *gray* vectors and *circles*.

Parameters of focal interest (e.g., brand, vector, and ideal point locations) can be better understood visually as is commonly done in most Multidimensional Scaling methods. Figure 2 shows the derived joint space for the two-dimensional, situation-level structural heterogeneity MDS model. The derived two-dimensional joint space map shows that the five brands are distributed throughout the four quadrants. Dimension I (horizontal) represents interactions with other medications where brands on the right-hand side have high interaction with a large class of other medications as opposed to those brands on the left-hand side. Dimension II (vertical) represents

TABLE 4.
Characteristics of the three groups of physicians in average number of prescription.

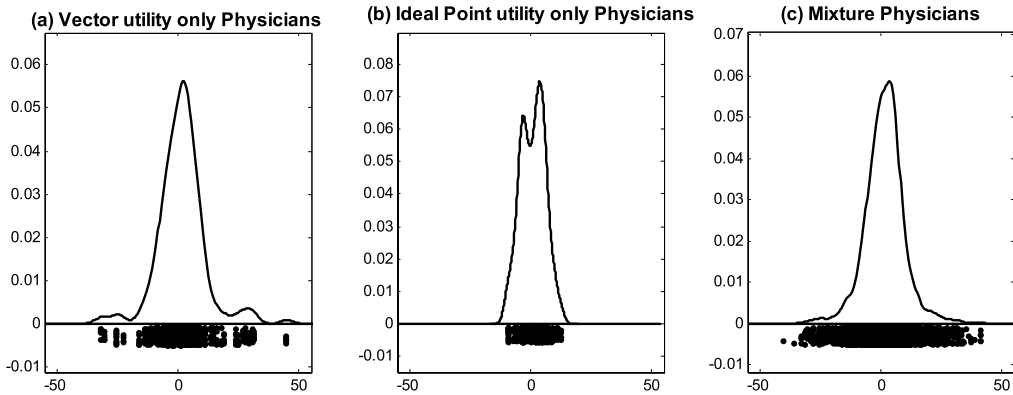|	Brand A	Brand B	Brand C	Brand D	Brand E	Overall
Vector utility only physicans	26.2	18.5	16.5	13.1	28.7	19.2
Ideal point utility only physicans	9.4	10.8	12.5	13.7	11.0	10.5
Mixture physicans	20.4	20.0	15.8	16.3	21.8	17.3

FIGURE 3.
Distribution of mean-centered prescriptions across the three groups of physicians.

age or order of entry of the brand into this market. In particular, brands on the top of this figure are older brands, whereas those located at the bottom of this dimension are newer brands.

Recall that an individual can have both vector and ideal point representations in the situation-level structural heterogeneity model (Model 2). In Figure 2, individuals whose observations are classified only to the vector model are represented with solid black vector termini, individuals whose observations are classified only to the ideal-point model are represented as black circles, while individuals with mixtures of vector and ideal point representations are depicted as gray vectors and circles. Note that 33 physicians (13.2%) comprise individuals with vector-only utility, and 13 physicians (5.2%) comprise individuals with ideal point-only utility. The clear majority of physicians (81.6%) turn out to have *both* vector and ideal point representations. Table 4 illustrates how these three groups of physicians can be best distinguished from each other by an examination of average prescription rates by brand (as well as overall) for each of the three derived groups. First, vector-utility-only physicians show the highest average prescription rates across brands, followed by physicians with mixture of vector and ideal-point utility. Physicians with ideal point-only utility show the lowest average prescription rates. Thus, physicians with ideal point-only utility can be described as low-volume physicians, while physicians with vector-only utility can be described as high-volume physicians. As shown in Table 4 vector-only utility physicians' number of prescriptions is almost two times higher than that of ideal point-only utility physicians in most brands.

A second distinguishing aspect concerning these group differences can also be gleaned from Figures 3 and 4. From Figures 3 and 4, it appears that the largest mixture group displays the most variation in overall prescription rates as compared to the other two groups, followed closely by the vector-only group. Notice how the physicians designated as the ideal point-only group have a much tighter concentration of prescriptions when viewing the distribution of either the raw or doubly centered prescriptions in these two figures. This contrast of variation also extends to characterizing the differences in prescriptions across the five brands. As aptly portrayed in
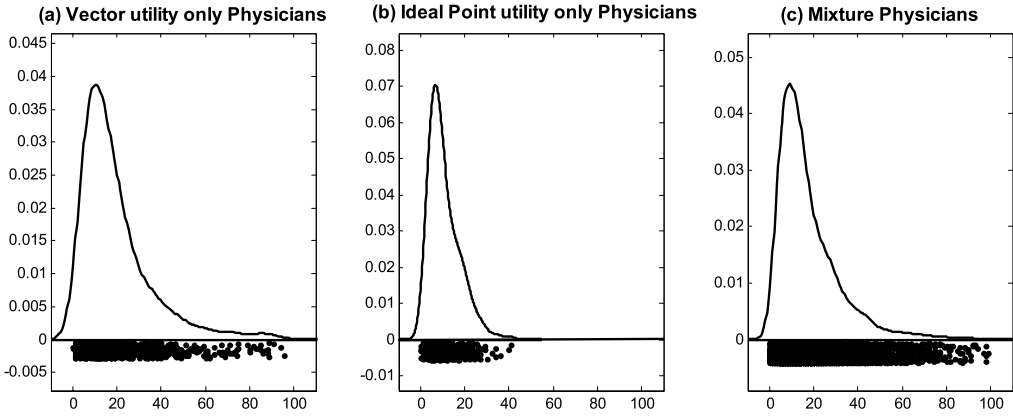
FIGURE 4.
Distribution of prescription volume across the three groups of physicians.

TABLE 5.
Structural change in prescription behavior for an illustrative physician.

(a) Number of prescription

Brand A	Brand B	Brand C	Brand D	Brand E
17	22	**30**	19	16
18	18	**24**	9	25
18	25	**23**	11	19
24	28	10	17	19
24	24	14	13	22
23	24	13	12	26

(b) Mean-centered data

Brand A	Brand B	Brand C	Brand D	Brand E
−5.81	0.02	12.69	1.53	−8.43
−2.81	−1.98	8.69	−6.47	2.57
−3.21	4.62	7.29	−4.87	−3.83
2.39	7.22	−6.11	0.73	−4.23
2.59	3.42	−1.91	−3.07	−1.03
1.39	3.22	−3.11	−4.27	2.77

Table 4, the vector-only group shows considerably more variation in prescription rates across all five brands of antidepressants compared to the other two groups. The ideal point-only group displays nearly the same prescription rates across the five brands.

Another aspect of the model solution illustrates that if there is any structural change in prescription behavior during the first six-month period, this change is captured by the structural heterogeneity. As an illustration, Table 5 shows the prescription behavior (Panel (a)) and corresponding mean-centered data (Panel (b)) for one specific physician. For the first three months, this physician highly prescribes Brand C, but s/he has a relatively lower number of prescriptions for this brand in the last three occasions. Notice that this pattern is reversed for Brand A. This change of prescription behavior is now represented as the change of signs for brands A and C in the mean-centered data (Panel (b)). It turns out that the situation-level structural heterogeneity model classifies the first three observations as ideal point structures, and the last three observa-
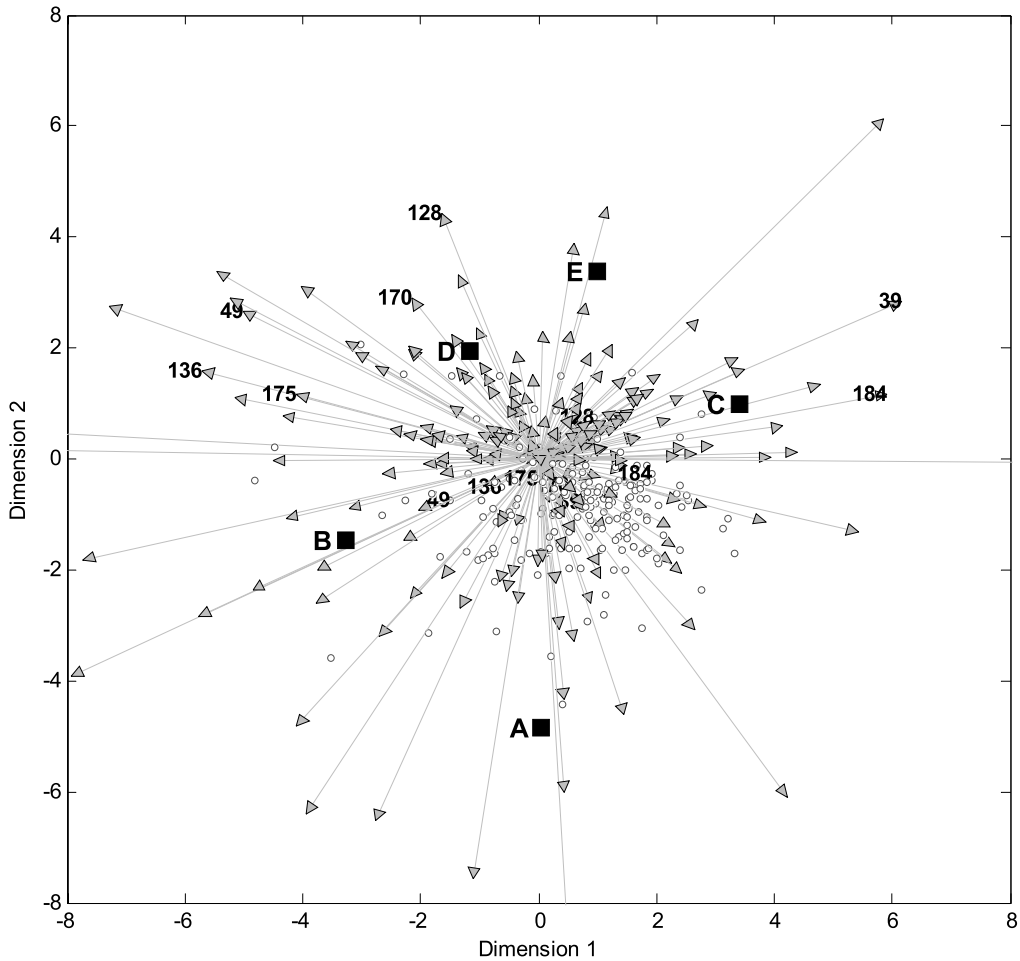
FIGURE 5.
Derived joint space for the mixture of vector and ideal point utility physicians. *Note*: only a subset of physicians' ideal points and vectors are labeled for illustration purposes. Ideal point and vector are labeled with the same number.

tions as vector structures. Figure 5 shows the derived two-dimensional joint space only for those physicians with mixtures of vector and ideal-point utility. In order to avoid confusion and reduce congestion, we label only an illustrative subset of the vector and ideal point values for a set of physicians; and, in order to better illustrate the contrasts between their respective vectors and ideal points, we placed them both on the same graph in Figure 5.

Here, the physician whose prescription behavior is described in Table 5 is labeled as 19. One can find his/her vector ($v = [-0.88, -0.47]$) in the third quadrant and his/her ideal point ($w = [1.12, -0.96]$) in the fourth quadrant. A quick inspection of these vector and ideal point coordinates shows that this ideal point has the closest distance to Brand C, and the vector shows that brands A and B have the highest preference. Similarly, one can find other physicians with mixtures of vectors and ideal points have their vectors and ideal points located in different quadrants. As such, one can infer that the major driver for situation-level structural heterogeneity is the change in prescription behavior over these time periods. Such changes in prescription behavior may be due to the associated difficulties of attempting to cure depression and the dynamic process physicians must go through in titrating various brands of antidepressants, changing brands due

TABLE 6.
Number of structural changes.

Number of times switching	Number of physicians	Percentage
0	59	24%
1	121	48%
2	48	19%
3	16	6%
4	4	2%
5	2	1%
	250	100%

Note: 0 indicates no mixtures of utility functions.

TABLE 7.
Number of situations used with vector model.

Number of situations for vector model	Number of physicians	Percentage
0	14	6%
1	23	9%
2	29	12%
3	39	16%
4	39	16%
5	61	24%
6	45	18%
	250	100%

Note: Number of observations for ideal point model is $6 -$ the number of observations for vector model with that revised quantity's associated frequencies and percentages.

to side effects experienced, experimenting with cocktails of multiple brands, trial-and-error prescription behavior of physicians in response to patient feedback on treating their symptoms, etc. Another possibility is that there could be an event that might change physicians' prescription behaviors in lieu of structural heterogeneity (e.g., an FTC warning about a particular drug that affects certain types of patients).

So what does the proposed analysis reveal—above and beyond standard MDS models—that is of managerial relevance? Table 6 illustrates how often physicians switch from one model to the other (i.e., vector to ideal point model or vice versa) over the six months. As one can see, a majority of physicians (48%) switched from one model to the other once. Table 7 shows the number of situations where physicians used a vector model (note, the number of situations used for the ideal-point model is just $6 -$ the number of observations used for the vector model, and one would look up the corresponding number of physicians and percentages for that quantity in this same table). Here, we see that the modal category appears to be five periods for which a vector model is used. Finally, Table 8 demonstrates the managerial usefulness of the approach which is not available in other MDS procedures for this type of data. We present a sample of some 25 physicians of the estimated indicator function χ_{ir} from (7) which indicates the particular type of response model employed by each physician for each situation/time period. This type of information can be gainfully employed by pharmaceutical marketers in direct-physician cam-

TABLE 8.
An illustrative sample of 25 physicians.

ID\replication	1	2	3	4	5	6
1	0	0	0	0	0	1
3	1	1	1	0	0	0
8	1	1	1	0	0	0
9	0	0	0	1	1	1
19	0	0	0	1	1	1
37	0	0	0	1	0	0
38	1	1	1	1	1	0
47	0	0	1	1	1	1
54	0	1	0	1	0	1
55	0	0	0	1	1	1
56	0	1	1	1	1	1
63	0	0	0	0	0	1
69	0	0	1	1	0	0
71	0	0	1	1	1	1
88	1	1	0	0	0	0
89	1	0	0	0	0	0
116	0	0	0	1	1	1
146	0	0	0	0	1	1
167	0	1	0	1	0	1
176	0	1	0	1	1	0
192	0	1	1	1	1	1
193	1	1	1	1	1	0
194	0	1	0	0	0	0
207	0	1	1	1	0	1
234	0	0	0	0	1	1

Note: 1 indicates vector utility and 0 indicates ideal point utility structures.

paigns which meet the specific needs of each physician, especially since syndicated data sources are available on prescription behavior for each and every physician in the US.

4. Discussion

Over the past two decades, numerous multidimensional scaling methods (MDS) have been developed to analyze dominance judgments. Two distinct types of MDS procedures (i.e., the vector and ideal point model) have been predominantly used to represent subjects' judgments. 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 such MDS models. We focused on the possibility that a sample of subjects may exhibit heterogeneous utility formulation and can be explained better by mixtures of the vector and the ideal point model. We introduced a new Bayesian MDS 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 accommodate preference heterogeneity at the individual level as well as at the situation level. We presented the details of the priors, likelihood, and posterior joint density, as well as the description of the full conditional distributions. An MCMC estimation procedure was presented as well.

We then applied the proposed approach to a pharmaceutical marketing application concerning doctors' prescriptions of the five leading brands of antidepressant medications. The

results demonstrate that a model that incorporates *both* structural and preference heterogeneity outperforms models without structural heterogeneity. This finding is quite consistent with existing literature that considers structural heterogeneity (Gilbride & Allenby, 2004; Jedidi & Kohli, 2005) in other model types. Our approach produces a parsimonious representation of preference structures via vector and ideal point representations. Furthermore, we can infer how a change in prescription behavior can be represented in the joint space map. No other MDS procedure exists to be able to perform this type of analysis.

Opportunities for future research merit discussion. Direct utilization of count data (e.g., the number of prescriptions) via a Poisson distribution or Negative Binomial distribution would be interesting, although extreme care needs to be given here due to problems of zero inflation (excessive zero prescriptions) as well as overdispersion. Also, we showed that a change in the prescription behavior is the major driver for situation-specific structural heterogeneity. The explicit incorporation of marketing activity (e.g., detailing) and background information (e.g., sociodemographic information and/or specialty) can aid in understanding why physicians show such structural heterogeneity. Finally, our current research is restricted by the assumption of a stationary joint space. Researchers have shown that the incorporation of nonstationary preference can enhance the understanding of subjects' behaviors (DeSarbo, Fong, Liechty, & Coupland, 2005; Liechty, Fong, & DeSarbo, 2005). Future research needs to incorporate nonstationary preference via a change-point models or dynamic state-space models, and the comparison with structural heterogeneity model with stationary space would be fruitful.

Appendix

Markov Chain Monte Carlo Algorithm for the Proposed Model

The estimation of the model parameters proceeds by recursively sampling from the following full conditional distributions. Because of limited space, we show only the full conditional distributions of Model 2.

1. Generate the individual structural heterogeneity indicator $I(\chi_{ir} = V)$ and $I(\chi_{ir} = IP)$. $I(\chi_{ir} = V) \sim \text{Bin}(1, P(\chi_{ir} = V | \sim))$ and $I(\chi_{ir} = IP) = 1 - I(\chi_{ir} = V)$, where the full conditional probability of structural heterogeneity parameter $P(\chi_{ir} = V | \sim)$ is computed by: $P(\chi_{ir} = V | \sim) = \frac{L_{ir|V} \phi_2}{L_{ir|V} \phi_2 + L_{ir|IP}(1-\phi_2)}$, and $P(\chi_{ir} = IP | \sim) = 1 - P(\chi_{ir} = V | \sim)$. Here, $P(\chi_{ir} = IP | \sim)$ indicates the full conditional distribution of the structural heterogeneity parameter given all other parameters.

2. Generate the vector model additive parameter $b_i | \tau_b^2, \chi_{ir}, x_{jt}, v_{it}$

$$P(b_i | \sim) \sim \begin{cases} N(\bar{b}_i, V_{b_i}) & \text{if } \sum_{r=1}^{R} I(\chi_{ir} = V) \geq 1, \\ N(0, \tau_b^2) & \text{otherwise,} \end{cases}$$

$$\text{where } V_{b_i} = \left[\sum_{r=1}^{R} \sum_{j=1}^{J} I(\chi_{ir} = V) + \frac{1}{\tau_b^2} \right]^{-1},$$

$$\bar{b}_i = \left[\sum_{r=1}^{R} \sum_{j=1}^{J} \left(\Delta_{ijr} - \sum_{t=1}^{T} x_{jt} v_{it} \right) I(\chi_{ir} = V) \right] V_{b_i}.$$

3. Generate the vector parameter $v_{it}|\tau_v^2, \chi_{ir}, x_{jt}, b_i$

$$P(v_{it}|\sim) \sim \begin{cases} N(\bar{v}_i, V_{v_i}) & \text{if } \sum_{r=1}^{R} I(\chi_{ir} = V) \geq 1, \\ N(0, \tau_v^2) & \text{otherwise,} \end{cases}$$

where $V_{v_i} = \left[\sum_{r=1}^{R} \sum_{j=1}^{J} x_{jt}^2 I(\chi_{ir} = V) + \frac{1}{\tau_v^2} \right]^{-1},$

$$\bar{v}_i = \left[\sum_{r=1}^{R} \sum_{j=1}^{J} \left(x_{jt} \left(\Delta_{ijr} - \sum_{l \neq t}^{T} x_{jl} v_{il} - b_i \right) I(\chi_{ir} = V) \right) \right] V_v.$$

4. Generate the ideal point model additive parameter $d_i|\tau_d^2, \chi_{ir}, x_{jt}, w_{it}, c_i$

$$P(d_i|\sim) \sim \begin{cases} N(\bar{d}_i, V_{d_i}) & \text{if } \sum_{r=1}^{R} I(\chi_{ir} = IP) \geq 1, \\ N(0, \tau_d^2) & \text{otherwise,} \end{cases}$$

where $V_{d_i} = \left(\sum_{r=1}^{R} \sum_{j=1}^{J} I(\chi_{ir} = IP) + \frac{1}{\tau_d^2} \right)^{-1},$

$$\bar{d}_i = \left(\sum_{r=1}^{R} \sum_{j=1}^{J} \left(\Delta_{ijr} - c_i \sum_{t=1}^{T} (x_{jt} - w_{it})^2 \right) I(\chi_{ir} = IP) \right) V_{d_i}.$$

5. Generate the ideal point model scale parameter $c_i|\tau_c^2, \chi_{ir}, x_{jt}, w_{it}, d_i$

$$P(c_i|\sim) \sim \begin{cases} N(\bar{c}_i, V_{c_i}) I(c_i < 0) & \text{if } \sum_{r=1}^{R} I(\chi_{ir} = IP) \geq 1, \\ N(0, \tau_c^2) I(c_i < 0) & \text{otherwise,} \end{cases}$$

where $V_{c_i} = \left(\sum_{r=1}^{R} \sum_{j=1}^{J} I(\chi_{ir} = IP) \left(\sum_{t=1}^{T} (x_{jt} - w_{it})^2 \right)^2 + \frac{1}{\tau_c^2} \right)^{-1},$

$$\bar{c}_i = \left(\sum_{r=1}^{R} \sum_{j=1}^{J} I(\chi_{ijr} = IP)(\Delta_{ijr} - d_i) \left(\sum_{t=1}^{T} (x_{jt} - w_{it})^2 \right) \right) V_{c_i}.$$

6. Generate the ideal point parameter $w_{it}|\tau_w^2, \chi_{ir}, x_{jt}, c_i, d_i$.

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

$$\alpha_w \left(w_{it}^{(n)}, w_{it}^{(o)} \right) = \min \left(\frac{P(w_{it}^{(n)})}{P(w_{it}^{(o)})}, 1 \right), \quad \text{and}$$

$$\frac{P(w_{it}^{(n)})}{P(w_{it}^{(o)})} = \frac{\prod_{r=1}^{R} [L_{ir|IP}(\Delta_{ijr}|w_{it}^{(n)}, rest)]^{I(\chi_{ir}=IP)}}{\prod_{r=1}^{R} [L_{ir|IP}(\Delta_{ijr}|w_{it}^{(o)}, rest)]^{I(\chi_{ir}=IP)}} \left(\frac{e^{-\frac{1}{2\tau_w^2}(w_{it}^{(n)})^2}}{e^{-\frac{1}{2\tau_w^2}(w_{it}^{(o)})^2}} \right),$$

where *rest* means other parameters in the likelihood.

7. Generate the brand parameter $x_{jt}|\tau_x^2, \chi_{ir}, v_{it}, b_i, w_{it}, c_i, d_i$.

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 $P(x_{jt(nc)}) \sim N(0, \tau_x^2)$ is used to generate the 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 $x_{jt(nc)}^{(n)} = x_{jt(nc)}^{(n)} + \omega e$, where ωe is a draw from a candidate generating density $N(O, \omega)$. Here, we calibrate ω needs so that the acceptance rate is around 30%, resulting in acceptable mixing probabilities as suggested by Gelman, Gilks, & Roberts (1996). Accept the new candidate $x_{jt(nc)}^{(n)}$ with probability

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

$$\frac{P(x_{jt(nc)}^{(n)})}{P(x_{jt(nc)}^{(o)})} = \frac{\prod_{i=1}^{I} L_i(\Delta_{ijr}|x_{jt(nc)}^{(n)}, rest)}{\prod_{i=1}^{I} L_i(\Delta_{ijr}|x_{jt(nc)}^{(o)}, rest)}\left(\frac{e^{-\frac{1}{2\tau_x^2}(x_{jt(nc)}^{(n)})^2}}{e^{-\frac{1}{2\tau_x^2}(x_{jt(nc)}^{(o)})^2}}\right),$$

where *rest* means other parameters in the likelihood.

Next, a random-walk Metropolis–Hastings algorithm with a Gamma prior $P(x_{jt(c)}) \sim G(sh_0, sc_0)$ and a Gamma proposal are 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 $k(x_{jt(c)}^{(o)})^2$ and the scale parameter as $\frac{1}{kx_{jt(c)}^{(o)}}$, so that the new candidate $x_{jt(c)}^{(n)}$ as the mean equal to the previous draw $x_{jt(c)}^{(o)}$ and the variance $\frac{1}{k}$ (Bradlow & Schmittlein, 2000). Therefore, a new candidate $x_{jt(c)}^{(n)}$ is generated from $G(k(x_{jt(c)}^{(o)})^2, \frac{1}{kx_{jt(c)}^{(o)}})$, and k needs to be tuned to get an adequate acceptance rate. Accept the new candidate $x_{jt(c)}^{(n)}$ with probability:

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

$$\frac{P(x_{jt(c)}^{(n)})}{P(x_{jt(c)}^{(o)}), 1} = \frac{\prod_{i=1}^{I} L_i(\Delta_{ijr}|x_{jt(c)}^{(n)}, rest)}{\prod_{i=1}^{I} L_i(\Delta_{ijr}|x_{jt(c)}^{(o)}, rest)}\left(\frac{\frac{1}{\Gamma(sh_2)sc_2^{sh_2}}(x_{jt(c)}^{(o)})^{sh_2-sh_0}e^{-x_{jt(c)}^{(o)}(\frac{1}{sc_2}-\frac{1}{sc_0})}}{\frac{1}{\Gamma(sh_1)sc_1^{sh_1}}(x_{jt(c)}^{(n)})^{sh_1-sh_0}e^{-x_{jt(c)}^{(n)}(\frac{1}{sc_1}-\frac{1}{sc_0})}}\right),$$

where $sh_1 = k\left(x_{jt(c)}^{(o)}\right)^2, sc_1 = \frac{1}{kx_{jt(c)}^{(o)}}, sh_2 = k\left(x_{jt(c)}^{(n)}\right)^2$, and $sc_2 = \frac{1}{kx_{jt(c)}^{(n)}}$.

8. Update the hyperparameter $\tau_x^{-2}|k_x, u_x, x_{jt}$

$$P(\tau_x^{-2}|\sim) \sim G\left(k_x + \frac{JT - 2T}{2}, \left(\frac{1}{2}\sum_{j=2T+1}^{J}\sum_{t=1}^{T}x_{jt}^2 + u_x^{-1}\right)^{-1}\right).$$

Similarly,

$$P(\tau_b^{-2}|\sim) \sim G\left(k_b + \frac{I}{2}, \left(\frac{1}{2}\sum_{i=1}^{I} b_i^2 + u_b^{-1}\right)^{-1}\right),$$

$$P(\tau_v^{-2}|\sim) \sim G\left(k_v + \frac{IT}{2}, \left(\frac{1}{2}\sum_{i=1}^{I}\sum_{t=1}^{T} v_{it}^2 + u_v^{-1}\right)^{-1}\right),$$

$$P(\tau_c^{-2}|\sim) \sim G\left(k_c + \frac{I}{2}, \left(\frac{1}{2}\sum_{i=1}^{I} c_i^2 + u_c^{-1}\right)^{-1}\right),$$

$$P(\tau_d^{-2}|\sim) \sim G\left(k_d + \frac{I}{2}, \left(\frac{1}{2}\sum_{i=1}^{I} d_i^2 + u_d^{-1}\right)^{-1}\right),$$

$$P(\tau_w^{-2}|\sim) \sim G\left(k_w + \frac{IT}{2}, \left(\frac{1}{2}\sum_{i=1}^{I}\sum_{t=1}^{T} w_{it}^2 + u_w^{-1}\right)^{-1}\right).$$

9. Update the hyperparameter $\phi_2|\chi_{ir}, a_2, b_2$

$$P(\phi_{-2}|\sim) \sim \text{Beta}\left(\sum_{i=1}^{I}\sum_{r=1}^{R} I(\chi_{ir} = V) + a_2, \sum_{i=1}^{I}\sum_{r=1}^{R} I(\chi_{ir} = IP) + b_2\right).$$

References

Belk, R.W. (1974). An Exploratory Assessment of Situational Effects in Buyer Behavior. *Journal of Marketing Research*, *11*, 156–163.

Belk, R.W. (1975). Situational Variables and Consumer Behavior. *Journal of Consumer Research*, *2*(3), 157–164.

Belk, R.W. (1979). A Free Response Approach to Developing Product-Specific Taxonomies. In A.D. Shocker (Ed.), *Analytical Approaches to Product and Marketing Planning*. Cambridge: Marketing Science Institute.

Berndt, E.R., Cockburn, I.M., & Griliches, Z. (1996). Pharmaceutical Innovations and Market Dynamics: Tracking Effects on Price Indexes for Antidepressant Drugs. In *Brookings Papers on Economic Activity, Microeconomics* (pp. 133–199). Brooking: Brookings Institution Press.

Bettman, J.R., Luce, M.F., & Payne, J.W. (1998). Constructive Consumer Choice Processes. *Journal of Consumer Research*, *25*(3), 187–217.

Bradlow, E.T., & Schmittlein, D.C. (2000). The Little Engines That Could: Modeling the Performance of World Wide Web Search Engines. *Marketing Science*, *19*(1), 43–62.

Busing, F.M.T.A., Groenen, P.J.K., & Heiser, W.J. (2005). Avoiding Degeneracy in Multidimensional Unfolding by Penalizing on the Coefficient of Variation. *Psychometrika*, *70*(1), 71–98.

Carroll, J.D. (1972). Individual Differences and Multidimensional Scaling. In R.N. Shepard, A.K. Romney, & S.B. Nerlove (Eds.), *Multidimensional Scaling; Theory and Applications in the Behavioral Sciences*. New York: Seminar Press.

Chib, S. (2002). Markov Chain Monte Carlo Methods. In S.J. Press (Ed.), *Subjective and Objective Bayesian Statistics* (2nd edn., pp. 119–171). New York: Wiley.

Consumer Reports. (2005). *Best Buy Drugs: Antidepressants*.

Coombs, C.H. (1964). *A Theory of Data*. New York: Wiley.

DeSarbo, W.S., & Carroll, J.D. (1985). Three-Way Metric Unfolding via Alternating Weighted Least Squares. *Psychometrika*, *50*(3), 275–300.

DeSarbo, W.S., & Cho, J. (1989). A Stochastic Multidimensional Scaling Vector Threshold Model for the Spatial Representation of Pick 'Any/N' Data. *Psychometrika*, *54*, 105–129.

DeSarbo, W.S., & Rao, V.R. (1984). GENFOLD2: A Set of Models and Algorithms for the GENeral UnFOLDing Analysis of Preference/Dominance Data. *Journal of Classification*, *2*, 147–168.

DeSarbo, W.S., & Rao, V.R. (1986). A Constrained Unfolding Methodology for Product Positioning. *Marketing Science*, *5*(1), 1–19.

DeSarbo, W.S., Manrai, A.K., & Manrai, L.A. (1994). Latent Class Multidimensional Scaling: A Review of Recent Developments in the Marketing and Psychometric Literature. In R.P. Bagozzi (Ed.), *Advanced Methods of Marketing Research* (pp. 190–222). Cambridge: Blackwell.

DeSarbo, W.S., Young, M.R., & Rangaswamy, A. (1997). A Parametric Multidimensional Unfolding Procedure for Incomplete Nonmetric Preference/Choice Set Data in Marketing Research. *Journal of Marketing Research*, *34*, 499–516.

DeSarbo, W.S., Kim, Y., Wedel, M., & Fong, D.K.H. (1998). A Bayesian Approach to the Spatial Representation of Market Structure from Consumer Choice Data. *European Journal of Operational Research*, *111*, 285–305.

DeSarbo, W.S., Kim, Y., & Fong, D. (1999). A Bayesian Multidimensional Scaling Procedure for the Spatial Analysis of Revealed Choice Data. *Journal of Econometrics*, *89*, 79–108.

DeSarbo, W.S., Fong, D.K.H., Liechty, J.C., & Coupland, J.C. (2005). Evolutionary Preferences/Utility Functions: A Dynamic Perspective. *Psychometrika*, *70*(1), 179–202.

Deun, K.V., Groenen, P.J.F., Heiser, W.J., Busing, F.M.T.A., & Delbeke, L. (2005). Interpreting Degenerate Solutions in Unfolding by Use of the Vector Model and the Compensatory Distance Model. *Psychometrika*, *70*(1), 45–69.

Diebolt, J., & Robert, C.P. (1994). Estimation of Finite Mixture Distributions through Bayesian Sampling. *Journal of the Royal Statistical Society. Series B (Methodological)*, *56*(2), 363–375.

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

Gelman, A., Gilks, W.R., & Roberts, G.O. (Eds.) (1996). *Efficient Metropolis Jumping Rules* (Vol. *5*). Oxford: Oxford University Press.

Gilbride, T.J., & Allenby, G.M. (2004). A Choice Model with Conjunctive, Disjunctive, and Compensatory Screening Rules. *Marketing Science*, *23*(3), 391–406.

Gilks, W.R., Richardson, S., & Spiegelhalter, D.J. (1996). Introducing Markov Chain Monte Carlo. In W.R. Gilks, S. Richardson, & D.J. Spiegelhalter (Eds.), *Markov Chain Monte Carlo in Practice* (pp. 1–19). London: Chapman & Hall.

Green, P.J. (1995). Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination. *Biometrika*, *82*(4), 711–732.

Harshman, R.A., & Lundy, M.E. (1984). Data preprocessing and the extended PARAFAC model. In H.G. Law & C.W. Snyder Jr. (Eds.), *Research Methods for Multimode Data Analysis* (pp. 216–284). New York: Praeger.

Harshman, R.A., & Lundy, M.E. (1985). The Preprocessing Controversy: An Exchange of Papers between Kroonenberg, Harshman and Lundy. University of Western Ontario, Department of Psychology.

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

Jedidi, K., & Kohli, R. (2005). Probabilistic Subset-Conjunctive Models for Heterogeneous Consumers. *Journal of Marketing Research*, *42*, 483–494.

Kamakura, W.A., Kim, B.D., & Lee, J. (1996). Modeling Preference and Structural Heterogeneity in Consumer Choice. *Marketing Science*, *15*(2), 152–172.

Karasu, T.B., Gelenberg, A., Merriam, A., & Wang, P. (2006). *Practice Guideline for the Treatment of Patients With Major Depressive Disorder*: The American Psychiatric Association.

Kass, R.E., & Raftery, A.E. (1995). Bayes Factors. *Journal of the American Statistical Association*, *90*(430), 773–795.

Liechty, J.C., Fong, D.K.H., & DeSarbo, W.S. (2005). Dynamic Models with Individual Level Heterogeneity: Applied to Evolution During Conjoint Studies. *Marketing Science*, *24*(2), 285–293.

Lopes, H.F. (2000). *Bayesian Analysis in Latent Factor and Longitudinal Models*. Durham: Duke Univ. Press.

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

Metropolis, M., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., & Teller, E. (1953). Equations of State Calculations by Fast Computing Machine. *Journal of Chemical Physics*, *21*, 1087–1091.

Newton, M.A., & Raftery, A.E. (1994). Approximate Bayesian Inference with the Weighted Likelihood Bootstrap. *Journal of the Royal Statistical Society. Series B (Methodological)*, *56*(1), 3–48.

Oh, M.S., & Raftery, A.E. (2001). Bayesian Multidimensional Scaling and Choice of Dimension. *Journal of the American Statistical Association*, *96*(455), 1031–1044.

Petty, R.E., & Cacioppo, J.T. (1986). *Communication and Persuasion: Central and Peripheral Routes to Attitude Change*. New York: Springer.

Richardson, S., & Green, P.J. (1997). On Bayesian Analysis of Mixtures with an Unknown Number of Components. *Journal of the Royal Statistical Society. Series B (Methodological)*, *59*(4), 731–792.

Rust, R., Simester, D., Brodie, R., & Nilakant , V. (1995). Model Selection Criteria: An Investigation of Relative Accuracy, Posterior Probabilities, and Combination of Criteria. *Management Science*, *41*(2), 322–333.

Slater, P. (1960). The Analysis of Personal Preference. *British Journal of Statistical Psychology*, *13*, 119–135.

Simon, H.A. (1955). A Behavioral Model of Rational Choice. *Quarterly Journal of Economics*, *69*(February), 99–118.

Simon, H.A. (1990). Invariance of Human Behavior. *Annual Review of Psychology*, *41*, 1–19.

Srivastava, R.K., Alpert, M.I., & Shocker, A.D. (1984). A Customer-Oriented Approach for Determining Market Structures. *Journal of Marketing*, *48*(2), 32–45.

Tanner, M.A., & Wong, W.H. (1987). The Calculation of Posterior Distributions by Data Augmentation. *Journal of the American Statistical Association*, *82*(398), 528–540.

Tucker, L.R. (1960). Intra-Individual and Inter-Individual Multidimensionality. In H. Gulliksen & S. Messick (Eds.), *Psychological Scaling: Theory and Applications* (pp. 155–167). New York: Wiley.

Tversky, A., & Kahneman, D. (1991). Loss Aversion in Riskless Choice: A Reference-Dependent Model. *Quarterly Journal of Economics*, *106*(November), 1039–1062.

Wedel, M., & DeSarbo, W.S. (1996). An Exponential-Family Multidimensional Scaling Mixture Methodology. *Journal of Business & Economic Statistics*, *14*(4), 447–459.

Wedel, M., & Kamakura, W. (2000). *Market Segmentation: Conceptual and Methodological Foundations*. Dordrecht: Kluwer Academic.

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