The Pennsylvania State University

The Graduate School

HIERARCHICAL BAYESIAN MODEL DEVELOPMENT WITH APPLICATIONS IN MARKETING

A Dissertation in

Statistics

by

Zhe Chen

Submitted in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

December 2014

The dissertation of Zhe Chen was reviewed and approved* by the following:

Duncan K. H. Fong
Professor of Marketing and Statistics
Dissertation Advisor, Chair of Committee

Steven F. Arnold
Professor of Statistics

Murali Haran
Associate Professor of Statistics

Ping Xu
Professor of Mathematics

Wayne S. DeSarbo
Smeal Distinguished Professor of Marketing

Aleksandra Slavkovic
Graduate Program Chair

*Signatures are on file in the Graduate School.

Abstract

Hierarchical Bayesian models have been widely used in practice to address different kinds of problems in many disciplines. Typically, efficient Markov Chain Monte Carlo simulation methods are developed to generate random samples from complicated posterior distributions resulting from such models. In the Marketing literature, many hierarchical Bayesian models have been devised to investigate various marketing phenomena. In this dissertation, I provide new hierarchical Bayesian models to investigate three marketing problems which have not been well studied in the literature, namely: (1) a heterogeneous Bayesian dynamic model to study the association between customer satisfaction and a firm's financial performance, (2) a Bayesian random-coefficient multinomial probit model to analyze customer choice panel data and evaluate the value of purchase history data in direct marketing, and (3) a Bayesian vector multidimensional scaling with variable selection procedure to produce a joint space map of consumers and brands for product positioning analysis. In addition to developing Bayesian models for these problems, we apply the methods to analyze real data and draw insights from our Bayesian analysis with marketing applications. Simulation studies with comparison results of some benchmark models will also be presented.

Table of Contents

List of Figures

List of Tables

Acknowledgments

It is a pleasure to thank the many people whose guidance, suggestions, help and support were greatly appreciated during my time as a graduate student in the past several years. I will never be able to finish this dissertation without these people.

My first earnest acknowledgement goes to my advisor, Dr. Duncan K. H. Fong, for his suggestive ideas and helpful guidance during these years. None of this work would have been possibly done without him. Not only did he teach me how to accomplish a project, he also encouraged me to think as a researcher. As a joint professor of Marketing and Statistics, he opened my eyes by introducing a different area of Statistics to me, which focus on the application of statistics in the real business world. I feel grateful that Duncan was always patient and supportive. He was willing to discuss the research ideas, suggest new directions of the projects, and give helpful advices and edits to improve my writings even if he was really busy.

Next, I would like to thank our helpful collaborator, Dr. Wayne S. DeSarbo. It is my great luck to have the opportunity to work with such a great researcher. He provided me valuable knowledge and suggestions for the marketing insights, the statistical methodologies, and new possibilities of the research projects. He helped me greatly in bridging the gap between the marketing and statistical problems in our research, and his passion to research encouraged my interests to this work. What's more, we share the same birthday! What a surprise!

I also wish to thank my thesis committee, Dr. Murali Haran, Dr. Steven F. Arnold, and Dr. Ping Xu. Dr. Haran was always willing to talk with me about my troubles with the research, and he was very considerate, supportive and friendly. Dr. Arnold also gave me suggestions and comments on my research. When I was serving as his TA, he gave me a lot of flexibilities, which earned me a lot of time to focus on my research. I would also like to thank both of these professors for their wonderful fundamental courses in Stochastic Process and Linear Models. In addition, I wish to thank Dr. Xu for kindly agreeing my request at the last moment.

I appreciate many kindnesses from the faculties in Statistic Department. Many thanks go to Dr. Runze Li. He was always there to talk and help whenever I have questions on academic research or personal life. He was more like a nice friend. I also wish to thank him for his generous financial support in my last summer so that I can focus on my research and the writing of this dissertation. I also want to thank Dr. Bruce Lindsay, Dr. Donald Richards, Dr. Naomi Altman, Dr. William Harkness and Dr. Dennis Lin for their helpful suggestions on my research and kind help when I had difficulty.

My most heartfelt thanks goes to my dear parents, who supported me with their unconditional love for every single and imaginable aspects during these years I am abroad. Especially, I want to thank my mom to be always talking with me in great patience. This helped me to relieve from the great pressure when writing the dissertation. I really want to express my deep appreciation and forever love to them.

Lastly, I also feel grateful to have so many great friends around me, Xianyun Mao, Qing Zhu, Yijia Feng, Jianping Sun, Tao Yang, Junyi Lin, Lina Ying, Li Xiao, Rui Wang. They provided me with their continuous encouragement and made my stay in State College happy and exciting. Appreciation also goes to Nancy Wang for her help on the writing of this dissertation. In addition, I want to thank my best friend Yunpeng Chen for his great friendship and maximum support. Special thanks go to my husband, Huajing Li. Were it not for his endless love, support, patience and tolerance, it would simply not have been possible for me to complete this work.

Chapter 1

Introduction

Bayesian analysis is an important tool for statistical inference and it is gaining popularity in the modern statistical literature and practice. Hierarchical Bayesian models are commonly employed in Bayesian analysis, where one assumes a hierarchical structure on the parameters of interest. In the past, such models were of limited use because of computational complexity. However, with the vast improvements in computing technology, this strategy has drawn more attention from both academicians as well as practitioners, especially in the past two decades. Researchers now make use of advanced simulation methodology (more specifically, Markov Chain Monte Carlo (MCMC) methods) to generate random deviates from complicated posterior distributions, and the current computing technology makes it possible that these simulation methods can be implemented in a practical and efficient manner.

Hierarchical Bayesian models possess an exceptional ability to provide estimates on parameters when the data availability is poor, while the comparable frequentist statistical techniques fail to achieve this. The modeling flexibility also allows the approach to be used to answer many different questions. Because of this adaptability, it has been widely used in various disciplines such as machine learning (Y. Zhang

& Koren, 2007), information retrieval (Blei, Jordan, & Ng, 2003), biology (Broet, Richardson, & Radvanyi, 2002), etc. In this dissertation, we focus on the development of Hierarchical Bayesian models applied to marketing research.

1.1 Background and Motivation

In the marketing domain, there is a plethora of literature using hierarchical Bayesian models to investigate various marketing phenomenon. For example, how to identify and understand respondents whose preference to a product is extreme so as to help improve product design and market segmentation (Allenby & Ginter, 1995); how to estimate individual-level-regression coefficients in cross-sectional data when there is only single observation per response unit (Fong, Ebbes, & DeSarbo, 2011); how to analyze data of repeated/sequential measurement to learn dynamic evolution of preference/utility (DeSarbo, Fong, Liechty, & Coupland, 2005; Liechty, Fong, & DeSarbo, 2005); how to deal with data from reduced experimental designs (Lenk, DeSarbo, Green, & Young, 1996), etc.

Here we propose Bayesian models to address three marketing problems that involve individual heterogeneity and/or time-varying characteristics. First, we study the relationship between customer satisfaction and a firm's shareholder value. The existent literature indicates that better customer satisfaction contributes to better financial performance (Anderson, Fornell, & Lehmann, 1994; Fornell, Johnson, Anderson, Cha, & Bryant, 1996); however, we argue that this contribution differs from firm to firm, and also changes over time. Second, we evaluate the effectiveness of micromarketing by analyzing a panel data set of consumer purchase choices. We build upon the work of Rossi, McCulloch, and Allenby (1996) and propose a Bayesian random-coefficient multinomial probit (RCMNP) model to provide estimates on each consumer's brand preference and the sensitivity to the marketing mix variables. Lastly, we investigate product positioning by proposing a Bayesian vector multidimensional scaling method with variable selection (BVMDS-VS) for the analysis of metric consumer rating data. Our method generates a joint space map of customers and brands and allows one to determine what brand attribute variables are significant in moving the relative positions of various brands on the map.

1.2 Challenges

There are several challenges of our research.

- The purpose of studying the customer satisfaction-shareholder value association is to construct a model that provides coefficients that accommodate for both subject heterogeneity and dynamic patterning. To explore the impact of demographic/firmographic variables on the association, one may expect the impact coefficient matrix to vary over time but we are not aware of any existing models that allow this;

- An existing RCMNP model to analyze panel data of household choices is sensitive to the prior specifications. For a special case when dealing with cross-sectional data, that is, only one observation per customer, it has been shown that the posterior distribution from that model is improper (Fong et al., 2011). Apparently we need a more robust procedure;

- Parameter identification is an important issue that needs to be solved in order to successfully implement Bayesian (vector) MDS models. One has to consider identification problems arising from rotation, reflection, permutation and expansion/shrinkage. If the identification problem is solved, then it is of interest to determine the significant factors that will affect the MDS solutions. However, we are not aware of any existing work that addresses these important problems together.

1.3 Contributions

From my perspective, this thesis makes the following contributions:

- We propose a heterogeneous Bayesian dynamic model to analyze replicated cross-sectional data that has been used to explore the relationship between customer satisfaction and shareholder value. Our model can give individual-level time-varying estimates even when there is only one observation per subject for each time period;

- A Bayesian random-coefficient multinomial probit model is proposed to analyze panel data of consumer choices. This model provides an improvement over an existing model by allowing more general assumptions and the resulting procedure is shown to be more robust to prior specifications. An efficient MCMC procedure is devised to compute the required Bayesian estimates;

- A Bayesian vector multidimensional scaling model with model selection procedure is developed to analyze consumer rating data which produces a joint space map of consumers and brands. The method can become an important tool in product positioning and development.

1.4 Thesis Organization

The remaining chapters of this thesis are organized as follows. In Chapter 2 we study the relationship between a firm's financial performance and its customer satisfaction index by using a new heterogeneous Bayesian dynamic model. Then, a general Bayesian random-coefficient multinomial probit model is presented in Chapter 3 to analyze panel data of consumer choices with an application in micro-marketing. In Chapter 4, we present a Bayesian vector multidimensional scaling method with model selection procedure to provide a joint space map of consumers and brands from customer rating data. In each of these chapters, we start by presenting the marketing motivation and statistical research problems, followed by a literature review on the related work. Then, we present our approach with a discussion on the model setting, technical challenges, prior specifications, and some technical details (the full conditional distributions and the proposed sampling strategy). Simulation study and an empirical data application are conducted and comparison results with a benchmark model will be reported. Finally, we conclude the contributions of the work and discuss future research directions at the end of each chapter. The dissertation concludes with a summary of the overall contributions and future research directions in Chapter 5.

Chapter 2

A Heterogeneous Bayesian Dynamic Model for Replicated Cross-sectional Data

2.1 Introduction

One of the fundamental findings of marketing theory is that customer satisfaction benefits firm performance (Anderson et al., 1994; Fornell et al., 1996). As the movement to adopt shareholder value-based measures of firm performance continues, there is a plethora of interest from marketing scholars in exploring the important influence of customer satisfaction on shareholder value-based firm performance (Anderson, Fornell, & Mazvancheryl, 2004; Gruca & Rego, 2005; Mittal, Anderson, Sayrak, & Tadikamalla, 2005; Fornell, Mithas, Morgeson, & Krishnan, 2006; Aksoy, Cooil, Christophe, Keiningham, & Yalcin, 2008). These research shows that customer satisfaction can create shareholder value by increasing future cash flow growth, and therefore indicates a positive association between customer satisfaction (SAT) and shareholder value (SV).

However, the effect of heterogeneity among firms has not been adequately studied in the literature. Individual firm heterogeneity indicates firms' differences and these differences often influence the effect of marketing strategies, polices, and decisions in diverse directions (Narasimhan & Zhang, 2000). To create shareholder value, firms adopt different customer strategies and different implementation of these customer strategies. As a result it is reflected in the capital market such that for different firms, shareholders would evaluate the effect of customer satisfaction differently. For example, although Yahoo is close to Google in customer satisfaction according to the University of Michigan's American Customer Satisfaction Index, Yahoo cannot match the fast growing of Google in terms of market growth, innovation, and stock price[1]. Therefore, the strengths of the association between customer satisfaction and shareholder value for these two firms are obviously different. Thus, it is crucial to incorporate heterogeneity in the modeling process and get firm-level association estimates to better understand the customer satisfaction effect for various firms.

In addition to heterogeneity, the usual assumption adopted in the literature that a positive association between customer satisfaction and shareholder value remains constant over time may merit more scrutiny. In order to cope with the changing environment and internal firm conditions, firms have to modify their strategies from time to time to satisfy both the financial environment and internal firm conditions. These different and time-varying strategies and implementation can create significant variations in performance across time, which are likely to result in a dynamic, rather than static, association between customer satisfaction and a firm's financial performance. The dynamic characteristic of the association can also be observed in the Google vs. Yahoo example. An inspection of our collected data suggestes that the magnitudes

[1] $http://www.forbes.com/2005/08/16/google-yahoo-satisfaction-cx_ld_0816google.html$

of the association for these two firms are different and they vary over time as well.

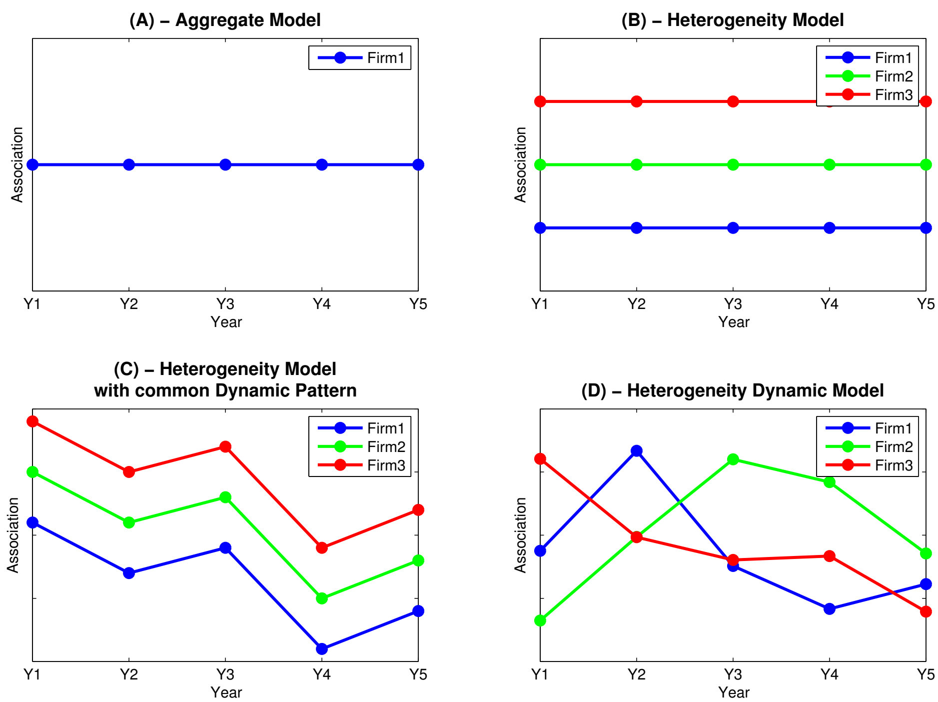
Indeed, researchers have just begun to realize the importance of heterogeneity in the effect of customer satisfaction on shareholder value and its dynamic characteristic. For example, using a hierarchical model and empirical Bayesian estimation method, Anderson et al. (2004) found a significant variation of magnitudes in the association between customer satisfaction and shareholder value at the industry level. While these authors have provided the *static* industry level estimates of such association, it is more desirable to obtain dynamic estimates at the individual firm level because not only that each firm is unique but the characteristics of a firm can also change over time. Some researchers (Graves, Kletter, Hetzel, & Bolton, 1998; Homburg, Koschate, & Hoyer, 2006) had investigated the dynamic nature of customer satisfaction, but they did not study the effect of customer satisfaction on shareholder value. Despite the strong recognition that customer satisfaction should be viewed from a dynamic perspective, the financial performance consequences of customer satisfaction from a dynamic perspective has not been systematically studied in the literature.

Although models with aggregate level parameters are commonly used in marketing research (e.g., Figure 2.1A), some elegant models have been introduced in conjoint analysis to uncover heterogeneity across firms and/or the dynamic nature of operational behavior(Figures 2.1B, 2.1C, and 2.1D). For example, Allenby and Ginter (1995) adopt a Bayesian random-effect model and apply it to a conjoint analysis of credit card attributes to study respondent heterogeneity. Lenk et al. (1996) generalize Allenby and Ginter's Model and analytically investigate the trade-off between the questionnaire length and the number of subjects involved that describe the part-worth heterogeneity. Bayesian random-effect models are successfully implemented in these two cases to give individual level, static, estimates of regression coefficients (Figure 2.1B). To model time dependent part-worths, Liechty et al. (2005) and DeSarbo et al. (2005) assume dynamic linear models (DLM) in their work. Furthermore, they develop the basic DLM to allow individual level regression coefficients. Liechty et al. (2005) show that the structure underlying preferences can change during the administration of repeated measurements and data collection due to learning, fatigue and boredom. Nevertheless, they only consider the situation when subjects follow a common dynamic pattern, as shown in Figure 2.1C. DeSarbo et al. (2005) use an advanced DLM to study the potential changes to the underlying prference/utility structure of the respondents in a psychological research. Their model gives individual level and time varying regression coefficients as shown in Figure 2.1D but it lacks flexibility in incorporating explanatory variables to help explain the variation on the individual-level dynamic regression coefficients. Mela et al. (1998) employ a varying-parameter model to investigate the long term effect of brand promotion on customer's stockpiling behavior. They provide customer-level time-varying coefficients to measure each individual's behavior, and use the existence of brand promotion to explain the varying parameters. In their model, they assume the impact of brand promotion is fixed as they intend to assess the long term effect. However, it is more general to assume the influence of the brand promotion to be changing over time and it is preferable to provide a temporal explanation of the impact in addition to dynamic coefficients.

So far, we have explained the need to develop a model in customer satisfaction research that allows for cross-sectional heterogeneity and temporal dynamics. In this chapter, we propose a heterogeneous Bayesian dynamic model which can be used to study the effect of customer satisfaction (SAV) on shareholder value (SV).

We present a framework that specifies how changing industry environment and firm characteristics would influence the SAT-SV association. Our research is among the first to investigate the important dynamic effects of firm heterogeneity and chang-

Figure 2.1. A diagram to show various scenarios of model settings



ing environments on the SAT-SV relationship by providing firm level estimates of the association over time. From our analysis, we find that industry environment and firm characteristics can dampen or amplify the individual firm level association between customer satisfaction and shareholder value. More importantly, these impacts are not static over time. We disentangle the dynamic complexity brought by individual firm heterogeneity and provide industry and firm level factors to help explain the complexity.

The remaining part of this chapter is organized as follows. In Section 2.2, we give a brief review of the existing models that are used to estimate heterogeneous and/or time dependent regression coefficients in the current literature. Section 2.3 presents our heterogeneous Bayesian dynamic model. We perform a simulation study to assess the performance of the proposed model and compare results with a benchmark model in Section 2.4. We then apply our model to a real data set to analyze the relationship between customer satisfaction and shareholder value in Section 2.5. After describing the data set and the measurement of our variables, we present the empirical results and findings obtained from our proposed model as well as those from the benchmark model to make comparisons. Section 2.6 gives a summary and concludes the chapter.

2.2 Literature Review

2.2.1 Bayesian Random-effect Model

Random-effect model is one kind of hierarchical linear model. It assumes that the coefficients in the regression setting are not fixed parameters, but rather random variables, which usually follows a specific distribution. With the vast development of computing technology in recent years, there is a plethora of literature using random-effect models to estimate regression coefficients by means of the Bayesian method (e.g., Bradlow, Wainer, & Wang, 1999; Yen, Liou, Lin, & Chen, 2006). In the marketing domain, Bayesian random-effect models are also adopted to explain marketing phenomenon and behavior (e.g., Rossi, Mcculloch, & Allenby, 1996; Rossi, Gilula, & Allenby, 2001).

Allenby and Ginter (1995) develop a Bayesian random-effect model and apply it to a conjoint study of credit card attributes. For ratings y_{it} from respondent i on profile t, $i = 1, \ldots, N$, $t = 1, \ldots, T$, their model assumes:

$$y_{it} = {x_{it}}'(\Delta + \beta_i) + \varepsilon_{it}, \quad \varepsilon_{it} \sim N(0, \sigma^2), \tag{2.1}$$

where Δ is a vector of fixed-effect coefficients that are constant across respondents, β_i is a vector of respondent-specific coefficients that modifies the fixed-effects, and x_{it}

is a vector of independent variables that are associated with respondent i on profile t. To describe how β_i differs among respondents through covariates z_i, a second level equation is employed:

$$\beta_i = \Gamma z_i + \zeta_i, \quad \zeta_i \sim N(0, D), \tag{2.2}$$

where Γ is a matrix of coefficients which relate β_i to z_i. Conjugate priors are then given on the parameters so that a Gibbs sampling algorithm (Gelfand & Smith, 1990) can be applied to obtain the parameter estimates.

Lenk et al. (1996) generalized Allenby and Ginter's model by introducing heterogeneity in the error variance. To be specific, instead of a common error variance, they assume that:

$$\varepsilon_{it} \sim N(0, {\sigma_i}^2). \tag{2.3}$$

All the other settings remain the same as in Allenby and Ginter (1995). The introduction of individual level variances brings computational complexity and a Metropolis algorithm (Hastings, 1970) is generally needed in the MCMC procedure to get estimates of the parameters. The consideration of individual level error variances contributes to the investigation on how the respondents may vary from each other on the measurement scale not only in location but also in spread.

The Bayesian random-effect model, although adequately describes the heterogeneity nature of the respondents and provides individual level coefficient estimates, it assumes that the relationship for each individual is all constant across time.

2.2.2 Bayesian Dynamic Linear Model (BDLM)

Bayesian Dynamic Linear Model (West & Harrison, 1997) is widely used to handle time series data where parameters of the later time periods are dependent on those of earlier time periods. Let Y_t be a vector of observations over time, $t = 1, \ldots, T$. The basic representation of BDLM is:

$$Y_t = F_t{}'\beta_t + v_t, \quad v_t \sim N(0, V_t); \tag{2.4}$$

$$\beta_t = G_t\beta_{t-1} + w_t, \quad w_t \sim N(0, W_t). \tag{2.5}$$

These two equations are called observation and evolution equations, respectively. Here,

- β_t is the time dependent part-worth, which is a parameter of interest;
- F_t is a design matrix with known values of independent variables;
- G_t is a evolution matrix with known covariates;
- v_t is observational error which is normally distributed with covariance matrix V_t;
- w_t is evolution error which is normally distributed with covariance matrix W_t;
- For each time period, both V_t and W_t are assumed known. In addition, observational error and evolution error are mutually independent.

There are various kinds of Bayesian dynamic linear model used in different disciplines of application, for example, finance and economics (Aguilar & West, 2000), medicine (Tonellato, 2001) and time series (Bhaumik, Dey, & Ravishanker, 2003).

The basic BDLM models the dynamic nature of some processes through use of the evolution equation. By introducing the evolution matrix G_t, the model relates the time dependent part-worths based on the assumption that the parameter of an earlier time period can help explain the current one. However, the basic model does not take heterogeneity into consideration so that it can only give aggregate estimates of the part-worths. To address this concern, some researchers have modified the basic dynamic linear model to incorporate heterogeneity in their work.

Liechty et al. (2005) have applied the Bayesian dynamic linear model on conjoint rating-based data. To investigate the association between some reference rating Y_{it} and independent variable X_{it} , the observation equation is defined as:

$$y_{it} = {x_{it}}'(\bar{\beta}_t + \beta_i) + v_{it}, \quad v_{it} \sim N(0, V_i). \tag{2.6}$$

Note the differences (and similarities) of Equation (2.6) from Equations (2.1) and (2.4). The fixed effect component across respondents in Equation (2.1) is allowed to vary with respect to time here, and this is equivalent to the dynamic part in BDLM (Equation (2.4)). By means of $\bar{\beta}_t$, we can track the changing effect of x_{it} on y_{it} across time. On the other hand, β_i provides insights on heterogeneity among respondents. In Equation (2.6), the observational variance is changing across individuals, but remains constant over time. This is different from the assumption of the basic dynamic linear model.

As usual, a second level equation is used to model the part-worths. For heterogeneity, the authors assume Equation (2.2) as in a random-effect model to relate β_i and z_i. For the time dependent part-worth $\bar{\beta}_t$, they use Equation (2.5) except that they drop the assumption of a known G_t, which is now assumed to be a random diagonal matrix. In addition, the authors proposed a second form of the evolution equation, which is given by:

$$\bar{\beta}_t = \bar{\beta}_0 + \bar{\beta}_1 t + \bar{\beta}_2 t^2, \tag{2.7}$$

where the coefficients in this dynamic functional form are assumed to have different normal priors.

Although one can obtain estimates of individual level part-worths β_i and aggregate time dependent part-worths $\bar{\beta}_t$ from this model, the estimate of β_{it} will take the restrictive form of $\bar{\beta}_t + \beta_i$ only. This is a limitation as the time dependent part-worths is not allowed to vary among respondents.

DeSarbo et al. (2005) use an improved dynamic linear model to analyze conjoint data. The assumption of a stable preference for all possible stimulus options as defined in normative theories of value maximization is challenged in that paper. The authors argue that preferences will change through a dynamic process due to learning, boredom, fatigue, etc. in the process of evaluation. The modified dynamic linear model adopts the following observation equation:

$$y_{it} = {x_{it}}'\beta_{it} + v_{it}, \quad v_{it} \sim N(0, V_i). \tag{2.8}$$

Note that this model does not treat the time dependent part-worth as an additive effect; instead, it estimates β_{it} for each respondent at every time period.

The evolution equation considered in DeSarbo et al. (2005) is:

$$\beta_{it} = \beta_{i,t-1} + w_{it}, \quad w_{it} \sim N(0, V_i W_{it}). \tag{2.9}$$

When $t = 0$, conditional on V_i,β_{i0} has a normal distribution with covariance matrix $V_i C_0$, where C_0 is pre-specified to be very "large", which implies vague prior information on β_{i0}. Conjugate prior distribution is assumed on V_i, and tuning variables W_{it} in the prior distributions are specified based on results from a training data set.

A limitation with the previous model is that the relationship assumed on the part-worth of consecutive time periods appears to be restrictive. Equation (2.9) implies that G_t is equal to the identity matrix for all t. It does not allow cases where the impact can be amplified or reduced during the dynamic process. A general shortcoming of the Bayesian Dynamic Linear Model is that, although it is able to provide individual-level, time varying estimates, it is difficult to include explanatory variables to explain such variation. Researchers will not be satisfied by merely knowing "things are different"; instead, they always want to know "why they are different".

2.2.3 Varying Parameter Model

Mela et al. (1998) developed a varying parameter model to analyze the long-term impact of promotions on consumer stockpiling behavior. To model consumer' purchasing incidence, they define a latent utility value, y_{it}, underlining the consumer i's purchase decision at time t and relate it to some explanatory variables x_{it} by a linear regression as follows:

$$y_{it} = {x_{it}}'\beta_{it} + \epsilon_{it}, \quad \epsilon_{it} \sim N(0, \sigma_0^2). \tag{2.10}$$

To evaluate the effect of marketing activities on consumer behavior, the authors construct a hierarchical structure to model the individual level, time-varying parameters β_{it}. Let z_{it} represent the long-term promotion indicator. Then, β_{it}, is reparameterized as a function of z_{it}, that is:

$$\beta_{it} = \Delta z_{it} + v_{it}, \quad v_{it} \sim N(0, D). \tag{2.11}$$

Note that Mela et al. (1998) assumes that D is a diagonal matrix so that each component of β_{it} is independent from each other. Although this setting is helpful for computational purposes, it is restrictive when comparing with a general positive definite matrix, which appears to be more realistic.

A second limitation of this model is that Δ in Equation (2.11) is assumed to be the same over time. For many problems it is more general to assume a dynamic impact matrix Δ_t instead of a static one.

Inspired by these three groups of models, we want to propose a model which contributes to the literature in the following aspects:

- The model can measure firm-level time-varying effects of customer satisfaction on shareholder value;
- Explanatory variables can be incorporated to provide an explanation for the varying parameters;
- It is possible to allow a dynamic impact matrix Δ_t to explain the variation in β_{it}.

We will present our proposed Heterogeneous Bayesian Dynamic Model that satisfies all of the above requirements in the next section.

2.3 The Proposed Model

2.3.1 Heterogeneous Bayesian Dynamic Model

For $i = 1, \ldots, N$ firms at $t = 1, \ldots, T$ time horizon, let Y_{it} be the log form of Tobin's q, the measure that we use to evaluate shareholder value. (We will explain it in details in Section 2.5.1.) We assume a regression setting of each Y_{it},

$$Y_{it} = X_{it}'\beta_{it} + \varepsilon_{it}, \tag{2.12}$$

where X_{it} is a 2×1 vector with one as the first element and the value of customer satisfaction index for firm i at time $t-1$ (to reflect a delayed effect) as the second element. β_{it} is a vector of regression parameters, and we allow individual level estimates for every firm at each time period. The error term ε_{it} is independent and normally distributed as $N(0, \sigma_t^2)$. If firmographic/demographic variables are available, they will be used in modeling the firm level regression coefficients,

$$\beta_{it} = \Delta_t Z_{it} + \delta_{it}, \tag{2.13}$$

where Z_{it} is a vector of dimension K containing one as the first element and the values of firmographic/demographic variables for firm i at time $t-1$ as the remaining elements. Δ_t is a $2 \times K$ matrix of the impact coefficients at time t, and δ_{it} is the error term which is independent and normally distributed as $N(0, \Sigma_t)$. We let $\Sigma_t = \sigma_t^2 C_t$ where C_t is a scale-free matrix and we note that such assumption is quite common in Bayesian dynamic linear modeling literature (cf., West & Harrison 1997).

2.3.2 Prior Specifications

To complete our model specification, we need to give prior distribution to C_t, σ_t^2, and Δ_t. For the first time period, we assume proper priors on all the parameters, namely, a Wishart prior distribution on the inverse of C_1, a gamma prior distribution on the inverse of σ_1^2, and a multivariate normal prior distribution on Δ_1. More specifically,

$$C_1^{-1} \sim W_J(aI_J, b), \tag{2.14}$$

$$\sigma_1^{-2} \sim \mathrm{Ga}(p, q), \tag{2.15}$$

$$\mathrm{vec}(\Delta_1) \sim N_{JK}(\eta_0, \gamma). \tag{2.16}$$

In Equations (2.15) and (2.16), we allow the variance to be arbitrarily large by sending p to 0 and q to infinity, and γ^{-1} to a zero matrix, respectively, to represent vague priors.

Note, caution should be taken in assigning non-informative and improper priors because the resulting posterior distribution can be improper (e.g., Berger, Oliveira, & Sanso, 2001). If an improper posterior distribution is derived due to an improper prior, it will highly affect our ability to correctly estimate the parameters. Fong, Ebbes and DeSarbo (2011) have proved that the model given by Equations (2.12) and (2.13), with priors specified in Equations (2.14), (2.15) and (2.16) (including the case of vague priors), will always result in a proper posterior distribution as long as some mild regularity conditions are met. They have also shown that a Wishart prior distribution on the inverse of Σ_1 (instead of C_1 in Equation (2.14) leads to an improper posterior distribution when the variances in Equations (2.15) and (2.16) are sent to infinity.

For subsequent time periods, we employ the posterior distribution of the parameters of the last time period to be the prior distribution of the parameters of the current time period. More specifically, let $D_{it} = \{Y_{it}, X_{it}, Z_{it}, D_{i(t-1)}\}$ be the information known at time t for firm i, and $D_{i0} = \emptyset$ (the null set). Then the prior distribution of $(C_t^{-1}, \sigma_t^2, \Delta_t)$ is:

$$\pi(C_t^{-1}, \sigma_t^{-2}, \Delta_t) = p_{t-1}(C_t^{-1}, \sigma_t^{-2}, \Delta_t | D_{i(t-1)}), \tag{2.17}$$

where p_{t-1} represents the posterior distribution of $C_{t-1}^{-1}, \sigma_{t-1}^2, \Delta_{t-1}$ at time $t-1$.

2.3.3 Full Conditional Distribution

In this section, we present the proposed Markov Chain Monte Carlo (MCMC) sampling algorithm to get estimates of the parameters $(\beta_{it}, C_t^{-1}, \sigma_t^2, \Delta_t)$ for $i = 1, \ldots, N$ and $t = 1, \ldots, T$, by first deriving the full conditional distributions of the parameters. When the full conditional distributions are all standard probability distributions, we can directly generate random estimates from such distributions, and the MCMC algorithm is called Gibbs sampling. When a full conditional distribution is not a standard probability distribution, we employ a Metropolis-Hasting algorithm by providing a proposal distribution from which random deviates will be chosen with certain probability of acceptance.

Random deviates are generated iteratively and recursively according to the following full conditional distributions (detailed derivation of these distributions is given in Appendix A).

- For $i = 1, \ldots, N$ and $t = 1, \ldots, T$, generate the coefficients β_{it} according to:

$$p(\beta_{it} | \text{ all others}) = N_J(\beta_{it}^0, \Psi_{it}), \tag{2.18}$$

  where

$$\Psi_{it} = \sigma_t^2 (X_{it} X_{it}' + C_t^{-1})^{-1},$$

$$\beta_{it}^0 = (X_{it}X_{it}' + C_t^{-1})^{-1}(Y_{it}X_{it} + C_t^{-1}\Delta_t Z_{it});$$

- For $t = 1, \ldots, T$, generate the $\mathrm{vec}(\Delta_t)$ according to:

$$p(\mathrm{vec}(\Delta_t)|\text{ all others}) = N_{JK}(\eta_t^0, \sigma_t^2\Theta_t^{-1}), \tag{2.19}$$

where

$$\begin{aligned}
\Theta_t =& \left[\gamma^{-1} + \sigma_t^{-2}\sum_{i=1}^{N}\left(\bar{Z}_{it}'C_t^{-1}\bar{Z}_{it} + \sum_{s=1}^{t-1}\frac{(X_{is}X_{is}' \otimes Z_{is}Z_{is}')}{X_{is}'C_tX_{is}+1}\right)\right]^{-1}, \\
\eta_t^0 =& \Theta_t^{-1}\left[\gamma^{-1}\eta_0 + \sigma_t^{-2}\sum_{i=1}^{N}\left(\bar{Z}_{it}'C_t^{-1}\beta_{it} + \sum_{s=1}^{t-1}\frac{Y_{is}X_{is}\otimes Z_{is}}{X_{is}'C_tX_{is}+1}\right)\right], \\
\bar{Z}_{it} =& I_J \otimes Z_{it}';
\end{aligned}$$

- For $t = 1, \ldots, T$, generate the $(\sigma_t^2)^{-1}$ according to:

$$p(\sigma_t^{-2}|\text{ all others}) = \mathrm{Ga}(\frac{1}{2}N(J+t) + p, Q_t^{-1}), \tag{2.20}$$

where

$$\begin{aligned}
Q_t =& \frac{1}{q} + \frac{1}{2}\sum_{i=1}^{N}\Bigg[(Y_{it} - X_{it}'\beta_{it})^2 \\
&+ (\beta_{it} - \Delta_t Z_{it})'C_t^{-1}(\beta_{it} - \Delta_t Z_{it}) + \sum_{s=1}^{t-1}\frac{(Y_{is} - X_{is}'\Delta_t Z_{is})^2}{X_{is}'C_tX_{is}+1}\Bigg];
\end{aligned}$$

- For $t = 1, \ldots, T$, the full conditional density of C_t^{-1} is:

$$p(C_t^{-1}|\text{ all others}) \tag{2.21}$$

$$
\begin{aligned}
&\propto |C_t|^{-\frac{N+b-J-1}{2}} \left[\prod_{i=1}^{N} \prod_{s=1}^{t-1} (X'_{is} C_t X_{is} + 1)^{-\frac{1}{2}} \right] \\
&\times \exp \Bigg\{ -\frac{1}{2} tr(a^{-1} C_t^{-1}) - \frac{1}{2\sigma_t^2} \sum_{i=1}^{N} \Bigg[(\beta_{it} - \Delta_t Z_{it})' C_t^{-1} (\beta_{it} - \Delta_t Z_{it}) \\
&\quad + \sum_{s=1}^{t-1} \frac{(Y_{is} - X'_{is} \Delta_t Z_{is})^2}{X'_{is} C_t X_{is} + 1} \Bigg] \Bigg\}.
\end{aligned}
$$

When $t = 1$, Equation (2.21) is proportional to the probability density function of

$$
W\left((a^{-1} I_J + \frac{1}{\sigma_1^2} \sum_{i=1}^{N} (\beta_{i1} - \Delta_1 Z_{i1})(\beta_{i1} - \Delta_1 Z_{i1})')^{-1}, N + b \right). \tag{2.22}
$$

Therefore, we will generate estimates from this Wishart distribution.

When $t > 1$, the full conditional distribution of C_t^{-1} does not follow a standard probability distribution and we do not know the explicit form of the normalizing constant. In this case, we use a Metropolis-Hasting algorithm to generate random deviates. The proposal function we apply is a Wishart distribution. For each iteration, we propose a new estimate C_t^{-1*} generated from the distribution:

$$
W\left((a^{-1} I_J + \frac{1}{\sigma_t^2} \sum_{i=1}^{N} (\beta_{it} - \Delta_t Z_{it})(\beta_{it} - \Delta_t Z_{it})')^{-1}, N + b \right). \tag{2.23}
$$

Then, we will accept the proposed estimate with probability:

$$
\begin{aligned}
\min \Bigg\{ 1, &\prod_{s=1}^{t-1} \prod_{i=1}^{N} \left(\frac{X'_{is} C_t^* X_{is} + 1}{X'_{is} C_t X_{is} + 1} \right)^{-\frac{1}{2}} \\
&\times \exp \left[\frac{\sigma_t^{-2}}{2} \sum_{i=1}^{N} \sum_{s=1}^{t-1} \left(\frac{(Y_{is} - X'_{is} \Delta_t Z_{is})^2}{X'_{is} C_t X_{is} + 1} - \frac{(Y_{is} - X'_{is} \Delta_t Z_{is})^2}{X'_{is} C_t^* X_{is} + 1} \right) \right] \Bigg\}.
\end{aligned} \tag{2.24}
$$

If the proposed estimate is not accepted, we will keep the current estimate of C_t^{-1}.

2.3.4 Model Comparison

Model comparison in Bayesian analysis is typically performed by comparing marginal likelihoods of various competing models. Based on our review of the dynamic models in the Marketing literature, we modify the model in Mela et al. (1998) which is then used as a benchmark model to compare with our proposed model. In the benchmark model, the diagonal covariance matrix of β_{it} is assumed to be a general positive definite matrix Σ. More specifically, Equation (2.11) becomes:

$$\beta_{it} = \Delta z_{it} + \nu_{it}, \quad \nu_{it} \sim N(0, \Sigma). \tag{2.25}$$

Conventional proper priors are then assumed for the parameters.

The marginal likelihood for model M_k, $Pr(\text{Data}|M_k)$, is given by:

$$\Pr(\text{Data}|M_k) = \int \Pr(\text{Data}|\Theta, M_k) f(\Theta|M_k) d\Theta,$$

where $\Pr(\text{Data}|\Theta, M_k)$ is the likelihood function given parameters Θ in model M_k, and $f(\Theta|M_k)$ is the prior density given M_k. Because of the integration required, exact computation of the marginal likelihood is often not possible. Various methods are available to estimate it, however. In this chapter two different methods are used: the harmonic mean and the Newton and Raftery's fourth estimator (see Appendix B for details). The harmonic mean estimator is easy to compute but can be problematic as it does not in general satisfy a Gaussian central limit theorem. The Newton and Raftery's fourth estimator is quite popular and it has been used by marketing researchers to calculate the marginal likelihoods (e.g., Gilbride, Allenby, & Brazell, 2006). Note that marginal likelihoods can be used to compute the Bayes factor, BF_{12}

which is given by:

$$BF_{12} = \frac{\Pr(M_1|(Data))/\Pr(M_1)}{\Pr(M_2|(Data))/\Pr(M_2)} = \frac{\Pr((Data)|M_1)}{\Pr((Data)|M_2)}$$

When BF_{12} is large, data are in support of M_1 relative to M_2. See Bolton, Fong, and Mosquin (2003) and Kass and Raftery (1995) for a comprehensive overview of the Bayes factor.

Another popularly used criterion to compare models is deviance information criterion (DIC) (Spiegelhalter, Best, Carlin, & Linde, 2002). Some scholars argue that Bayes factor typically favors a model that is more complex, so a certain penalty should be imposed on the complexity of the model when doing model comparison. Basically, DIC is formulated to penalize model both by the value of goodness-of-fit and the effective number of parameters. A model with smaller DIC value is preferred to a model with larger DIC. Details of the calculation of DIC are presented in Appendix B.

2.4 Simulated Data Example

A simulation study is performed to examine the performance of the proposed Heterogeneous Bayesian Dynamic model, and compare it with the benchmark model based on Mela et al. (1998). For simplicity, we denote our proposed model by HBDM and the benchmark model by MJB hereafter.

We randomly generate the observed covariates X and Z from a uniform distribution $U(-2, 2)$. To generate the true parameter values, we use the following scheme:

- When Year= 1, $C_1^{-1}, \sigma^{-2}, \text{and} \Delta_1$ are generated from their corresponding prior distributions in Equations (2.14), (2.15), and (2.16);

- When Year$= t > 1$, we add small values to the parameters in Year$= t-1$. For example, we add a small positive definite matrix to C_1^{-1} to obtain C_2^{-1}. Note that we make the difference of Δ_t between different years to be very small so that the generated data do not necessarily favor our proposed model.

Given $C_t^{-1}, \sigma_t^{-2}, \Delta_t$, we generate β_{it} according to Equation (2.13) and then Y_{it} according to Equation (2.12). Furthermore, in order to assess (1) whether or not Firm i's customer satisfaction in Year t will influence its shareholder value and (2) whether or not any specific firmographic/demographic variables of Firm i in Year t contribute to the SAT-SV association, we modify the values of Δ_t and β_{it} by setting the components with a magnitude smaller than 0.5 at 0.

To implement the proposed model, we adopt the following hyper parameters as defined in Equations (2.14), (2.15) and (2.16): $a = 1, b = K + 4, p = 3, q = 1, \eta_0 = 0$ and $\gamma = 100I$. Such hyper parameters are chosen such that the corresponding prior distributions represent vague information.

We first compute the Bayes factor (BF) to compare HBDM and MJB. We use two methods to calculate the log-marginal likelihood, the harmonic mean and the 4^{th} formula in Newton and Raftery (1994), to check the accuracy of the estimate. In addition, since there are more free parameters in HBDM than MJB, we use the Deviance Information Criterion (DIC) as another measure to compare models so as to validate the use of more free parameters. Table 2.1 lists the log-marginal likelihoods and their corresponding log BF of each model, as well as the DIC values. Following the guideline in Kass and Raftery (1995) and Spiegelhalter et al. (2002), both BF and DIC provide strong evidence in favor of our proposed model.

Next, we check the power of our proposed model to recover the true parameter values in terms of mean absolute deviation (MAD) and root mean squared error

Table 2.1. Simulation Analysis - Model Comparison Results.

Measurement	HBDM	MJB
Log marginal (Harmonic Mean)	-994.77	-1111.90
Log Bayes Factor	0.00	117.13
Log marginal (Newton and Raftery)	-975.74	-1108.90
Log Bayes Factor	0.000	133.16
DIC	1962.50	2225.80

(RMSE), shown in Table 2.2. Comparative measurements based on MJB are also presented. For all the parameters of interest, HBDM outperforms MJB in both MAD and RMSE. Note that the improvement in Δ of HBDM over MJB is small because we intentionally control the difference on Δ between years to be small so that the constant Δ assumption of MJB is approximately valid.

Table 2.2. Simulation Analysis - Parameter Recovery.

Variable	MAD		RMSE	
	HBDM	MJB	HBDM	MJB
β	1.239	1.617	1.575	2.079
σ	0.559	1.422	0.559	1.422
Δ	0.390	0.426	0.457	0.516
Σ	0.740	1.536	1.073	2.577

Lastly, we investigate how well each model can identify the significant parameters. For each component of β_{it} and Δ_t, we compute the corresponding probability of that parameter being positive (denoted by PP). At a 0.1 critical level, we tag the component to be positive if PP is greater than 0.9, negative if PP is less than 0.1 and zero otherwise. Then, by comparing the sign identification from the model with the sign of the true parameters, we compute the percentage of correct sign identification. Comparable results from both models are presented in Table 2.3. It is clearly shown that our proposed model pertains a higher accuracy than the benchmark model for both β and Δ.

Table 2.3. Simulation Analysis - Accuracy on Identifying the Significant Parameters.

Variable	HBDM	MJB
β	0.926	0.896
Δ	0.756	0.711

To summarize, based on this simulation study, we conclude that our model outperforms the benchmark model in three aspects: (1) The BF indicates that our model has a better fit than the benchmark model, and DIC further supports this argument after adjusting the effect of more free parameters in the model, (2) the proposed model can provide better estimates to the true parameter value in terms of MAD and RMSE, and (3) more accurate identification on the signs of the significant regression coefficients and impact matrix elements are achieved by the proposed model as compared to the benchmark model.

2.5 Application: Customer Satisfaction Data

2.5.1 Data and Measures

We collected a longitudinal data set from multiple archival sources, and applied our proposed Bayesian dynamic heterogeneity model to investigate the relationship between customer satisfaction and shareholder value.

First of all, we use Tobin's q to operationalize shareholder value (SV). Tobin's q is derived from stock market price, reflecting future performance and the market valuation of the firm; in other words, it is a forward-looking measure. Furthermore, Tobin's q can be used for comparisons across industries because it is not affected by accounting conventions. We employ Chung and Pruitt's (1994) method to compute

Tobin's q, which is listed in Table 2.4. We use COMPUSTAT, a database of marketing information on companies all around the world, to collect data from 1999 to 2007 to compute Tobin's q. In addition, similar to previous studies (e.g., Anderson et al. 2004), we take logarithm on Tobin's q, denoted by lnQ, to eliminate any possible violation of normality for the regression part of our model.

Secondly, to measure customer satisfaction (SAT), we selected companies included in the American Customer Satisfaction Index (ACSI) database. This satisfaction database has been successfully employed by a growing body of marketing researchers (Aksoy et al., 2008; Anderson et al., 2004; Morgan & Rego, 2006), and it is a measurement ranged from 0 to 100. Because we believe that customer satisfaction has a lagged effect on shareholder value, we only focus on data from 1998 to 2006. To be consistent with previous studies (e.g., Morgan & Rego, 2006), we removed utilities firms as well as privately held companies from our data. The final data set contains 630 observations, representing 70 different firms over 9 consecutive years.

Regarding inter-industry firm heterogeneity, we use two measures to capture the firm diversity from industry dynamism: the Hirschmann-Herfindahl index (HHI) and demand growth (DG). HHI is the sum of the square of all suppliers' market shares in an industry. This measure is a widely used indicator of market competitive structure (Anderson et al., 2004; Morgan & Rego, 2006). We used COMPUSTAT to compute HHI values for each of the industries categorized by the 4-digit SIC code in our data set for each year and scaled these values to be between zero (competitive) and one (concentrated). Also, we computed the average 12-month growth in industry sales to control for the differing industry demand conditions facing the firms in each of the four years. DG is computed as the ratio of industry sales increase versus the total industry sales.

There are several firm factors that may be relevant, namely, firms' relative advertising (AD), Research & Development intensity (RD) and firm assets (AS). Advertising intensity and R&D intensity are positively associated with firms' ability to bring superior products to market, and thus they are believed to be related to firm performance. Also, Asset is used to control for any scale economies that may impact firm performance (Morgan & Rego, 2006). Using data from COMPUSTAT for the 70 firms from 1998 to 2006, we calculate the control variables given in Table 2.4. Similar to customer satisfaction, one-year lagged variables are used in our model. The descriptive statistics for each of the variables in our data set for each of the nine years are presented in Table 2.5.

Table 2.4. Empirical Analysis: Data Measurements.

Variables	Measurements
Shareholder Value	$\text{Tobin's q} = \dfrac{\text{MVE} + \text{PS} + \text{DEBT}}{\text{TA}}$
Customer Satisfaction	ACSI index
AD Intensity	$\dfrac{\text{Advertising Expense}}{\text{Sales}}$
RD Intensity	$\dfrac{\text{R\&D Expense}}{\text{Sales}}$
Asset	Total Assets
Demand Growth	$\dfrac{\text{Industry Sales}_t - \text{Industry Sales}_{t-1}}{\text{Industry Sales}_t}$
Market Concentration	Hirschmann-Herfindahl Index (HHI)

where MVE = (closing price of shares at the end of the financial year) $\times$ (number of common shares outstanding); PS = liquidation value of the firm's outstanding preferred stock; DEBT = (current liabilities - current assets) + (book value of inventories) + (long-term debt); TA = book value of total assets.

Table 2.5. Empirical Analysis: Descriptive Statistics of Variables.

Variable		1999	2000	2001	2002	2003	2004	2005	2006	2007
ln(Q)	Mean	1.89	1.65	1.50	1.32	1.41	1.42	1.34	1.50	1.51
	S.D.	1.59	1.35	1.16	1.01	1.00	0.98	0.93	1.21	1.59
SAT	Mean	75.64	76.69	76.37	77.00	77.33	77.27	76.89	77.83	77.43
	S.D.	6.88	6.97	7.07	6.24	6.09	6.24	7.04	6.84	7.38
AD	Mean	0.05	0.04	0.04	0.04	0.04	0.04	0.04	0.05	0.05
	S.D.	0.05	0.04	0.04	0.04	0.04	0.04	0.05	0.07	0.10
RD	Mean	0.05	0.05	0.06	0.07	0.07	0.07	0.06	0.07	0.10
	S.D.	0.08	0.07	0.07	0.10	0.10	0.10	0.08	0.10	0.21
AS	Mean	9.76	9.83	9.92	9.96	10.01	10.06	10.11	10.19	10.29
	S.D.	1.45	1.47	1.46	1.46	1.49	1.55	1.54	1.55	1.54
DG	Mean	0.07	0.02	-0.01	0.01	0.03	0.05	-0.03	0.07	-0.05
	S.D.	0.08	0.12	0.09	0.16	0.26	0.22	0.14	0.12	0.16
HHI	Mean	0.20	0.20	0.21	0.22	0.23	0.23	0.24	0.24	0.28
	S.D.	0.16	0.16	0.16	0.16	0.18	0.19	0.21	0.21	0.23

2.5.2 Results and Findings

We applied our proposed Heterogeneous Bayesian Dynamic Model as well as the benchmark model to the data as described in the previous section. To faciliate a direct comparison, we use an informative prior on Year 1 for our proposed model by computing the tuning variables based on the random draws from the posterior distribution of the benchmark model. Specifically, let $\sigma^{-2} = \Big((\sigma^{-2})^{(1)}, \ldots, (\sigma^{-2})^{(R)}\Big)$, $\Delta = \Big(\Delta^{(1)}, \ldots, \Delta^{(R)}\Big)$ and $\Sigma^{-1} = \Big((\Sigma^{-1})^{(1)}, \ldots, (\Sigma^{-1})^{(R)}\Big)$ be the MCMC draws from MJB, where R denotes the total number of iterations. Then we set,

- $q = \text{var}(\sigma^{-2})/E(\sigma^{-2})$ and $p = E(\sigma^{-2})/q$;
- $\eta_0 = E(\Delta)$ and $\gamma = \text{cov}(\Delta)$;
- $a = 1$ and $b = \text{cov}(\Delta)$.

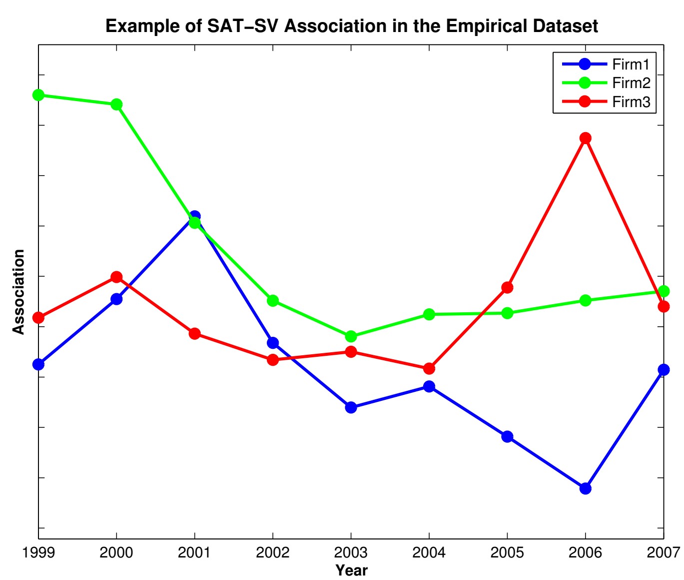
First of all, we compare our proposed model and the benchmark model in terms of Bayes factor and DIC. Table 2.6 presents the results for both models. Clearly our proposed model outperforms the benchmark model on both measures.

Table 2.6. Empirical Analysis - Bayes Factor and DIC.

Measurement	HBDM	MJB
Log marginal (Harmonic Mean)	-526.52	-598.42
Log Bayes Factor	0.0	71.90
Log marginal (Newton and Raftery)	-492.60	-558.05
Log Bayes Factor	0.00	65.45
DIC	1168.96	1191.56

Recall that Figure 2.1 depicts 4 scenarios that have been studied in the marketing research literature and it is of interest to know which pattern is most appropriate for our data. Before reviewing the results on all 70 firms, we provide the SAT-SV association over 9 years for the first three firms in our data set. Figure 2.2 is a line chart of the posterior means for these three firms. It is similar to Figure 2.1D in that the estimates are changing over time for these firms and the patterns are not identical. Therefore, a heterogeneity model with time-varying parameter looks reasonable.

Figure 2.2. Bayes Estimates (Posterior Means) for Three Firms in Our Dataset.



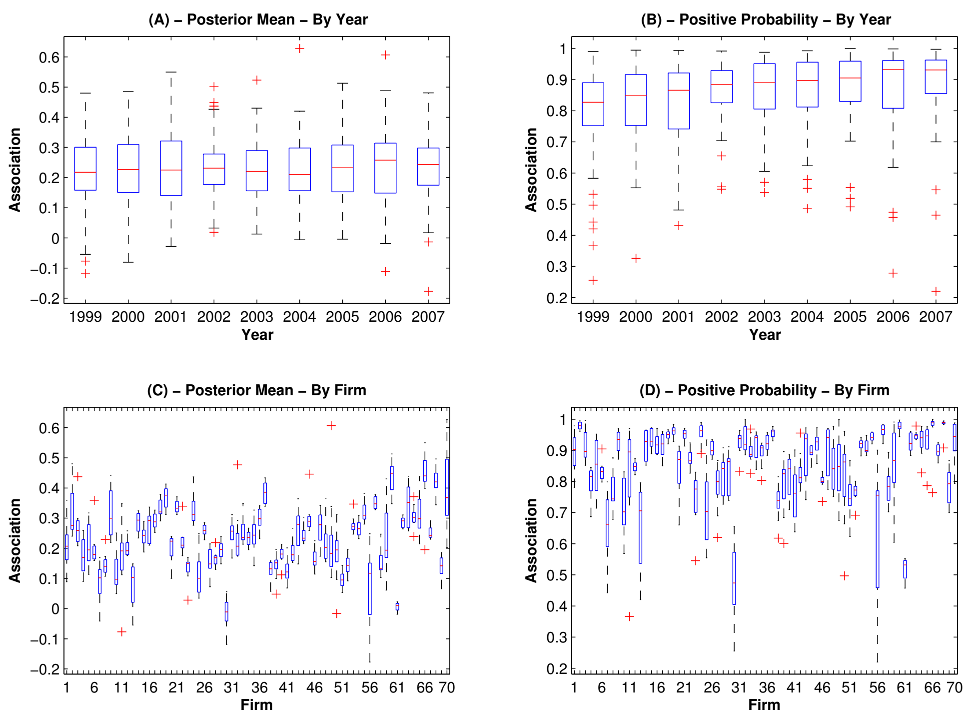
We have argued that the association between customer satisfaction and firms' shareholder value not only differs from firm to firm but also has a temporal variation. Figure 2.3 shows four boxplots of the SAT coefficients that are created for each year over all 70 firms and for each firm over all 9 years. We consider two measurements here: the posterior mean as well as the probability of that component being positive. Following Rossi et al. (1996), we use 0.9 as the threshold value to determine significance here. From Figure 2.3 we observe that:

- *The SAT-SV association has substantial heterogeneity among firms.* For some of the firms, there are even negative posterior estimates for the SAT coefficient (although not significant), while the majority of the point estimates are positive, as shown in Figure 2.3A. Figure 2.3B shows that, in each year, only part of the firms have significant SAT-SV association. When we check the boxplots of the posterior mean for each firm in Figure 2.3C, the divergence among firms is more pronounced. On the one hand, the gap between the largest median and the smallest median of the posterior means is not small. Firm 60 has a median as high as 0.45 while Firm 30 only gets -0.01, which suggests that the average level of SAT-SV association for each firm is not quite the same. On the other hand, the fluctuation of such association for each firm is also varying, as indicated by the different range of those boxplots. Some firms change dramatically from year to year, for example, Firm 49; some firms remain in a stable level for all 9 years, an example of which is Firm 61. Furthermore, Figure 2.3D shows some firms always have significant SAT-SV association, while some have minimum or even no effect. These findings indicate that it is not appropriate to assume an aggregate model for this data set;

- *The SAT-SV association is indeed changing over time for each firm.* Over the

70 firms in our data set, 10 firms have straight significant association for all years and 18 firms do not have significant association for any year. The other 60% of the firms have significant relationship for some years but non-significant relationship for the other years. Thus, if we adopt a static model to study the SAT-SV association, we will miss all these meaningful information and arrive at a wrong conclusion.

Figure 2.3. Bayes Estimates of the SAT Coefficient for the 70 Firms.



One of the major differences of our proposed model from the benchmark model is that our model uses the second level of the hierarchical setting to **dynamically** explain the variation of the SAT-SV association caused by individual firm heterogeneity. The posterior results for each component of the impact matrix that help explain the SAT-SV association are summarized in Table 7 for both models. The three rows for each parameter in each year represent the posterior mean and the standardized posterior mean of the SAT coefficient and the corresponding posterior probability for that parameter to be positive, respectively. In the last column, we present the aggregate level estimates obtained from the competing MJB (benchmark) model. The results show that generally HHI and AS contribute positively to the SAT coefficient. In other words, when a firm's competitive industry environment becomes more intense and the firm size is large, there is a closer tie between SAT and SV. Although the two models match on the inference of AS and HHI, they differ on other variables. MJB indicates that an increase in AD is associated with a decrease in SAT-VS relationship, but our proposed model shows that this negative effect does not always hold true. For some years, AD simply has no effect at all, from our model. In addition, MJB implies that DG is **not** a useful explanatory variable but our model finds that DG actually has a negative significant effect for some years. In other words, the association is strengthened for firms in industries with low demand growth for some years, based on our model. These important information will be missed if a static impact matrix is assumed in a model, as in the benchmark model.

In Bayesian MCMC computation, it is essential to check convergence of the generated Markov chain. Indeed there are many methods in the literature for MCMC convergence diagnostics (e.g., Brooks, 1998; Brooks & Roberts, 1998). In our study, we start the Markov chain with several different starting values, and inspect the trace plots to check whether the simulated chains will eventually visit the same region of the vector space. We also implement a standard CODA function which provides support for convergence (Cowles, 1996; Newton & Raftery, 1994). In addition, the between-chain and within-chain dispersions are compared by the Gelman and Rubin (1992) diagnostic. Our results show that convergence is achieved within 25000 iterations.

Table 2.7. Empirical Application: Posterior Means of some Δ Coefficients and the Corresponding Standardized Posterior Means and Probabilities that the Parameter is positive.

Variable	HBDM 1999	2000	2001	2002	2003	2004	2005	2006	2007	MJB
	0.23	0.23	0.23	0.23	0.23	0.22	0.23	0.24	0.23	0.23
Int	7.18	7.53	7.85	8.29	8.38	8.71	9.03	9.45	9.79	6.88
	1.00	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
	-0.04	-0.03	-0.03	-0.03	-0.03	-0.03	-0.03	-0.04	-0.04	-0.05
AD	-1.31	-1.01	-0.92	-0.86	-0.98	-1.20	-1.23	-1.42	-1.72	-1.30
	0.09	0.16	0.18	0.19	0.17	0.12	0.11	**0.08**	**0.04**	**0.01**
	0.01	0.01	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.00
RD	0.26	0.41	0.82	0.88	0.98	0.96	0.78	0.50	0.32	0.15
	0.60	0.66	0.79	0.81	0.83	0.83	0.79	0.69	0.63	0.55
	0.06	0.06	0.07	0.07	0.07	0.07	0.07	0.06	0.06	0.06
AS	1.73	1.86	2.26	2.33	2.51	2.55	2.45	2.41	2.25	1.50
	0.96	**0.97**	**0.99**	**0.99**	**0.99**	**0.99**	**0.99**	**0.99**	**0.99**	**0.93**
	-0.04	-0.04	-0.04	-0.04	-0.04	-0.03	-0.03	-0.04	-0.03	-0.03
DG	-1.06	-1.12	-1.23	-1.42	-1.38	-1.13	-1.17	-1.39	-1.17	-0.89
	0.14	0.13	0.11	**0.08**	**0.08**	0.13	0.12	**0.08**	0.12	0.20
	0.05	0.04	0.04	0.05	0.05	0.06	0.05	0.05	0.05	0.05
HHI	1.42	1.37	1.46	1.69	1.87	2.03	1.99	1.93	1.97	1.45
	0.92	**0.92**	**0.92**	**0.96**	**0.97**	**0.98**	**0.98**	**0.97**	**0.98**	**0.92**

Note: **Bold** indicates probability is over 0.90 or below 0.10.

2.6 Conclusion

In this chapter, we propose a heterogeneous Bayesian dynamic model to investigate the relationship between customer satisfaction and firm financial performance. The model provides firm level estimates of the association, which indicate that individual firm differences contribute to the large variation in magnitude of the association over time. Moreover, our model helps to identify the temporal effects of customer satisfaction in the way that it affects a firm's financial performance over time.

By employing the joint posterior distribution from the previous time period as the prior distribution of the current time period, our model takes a dynamic perspective which improves upon the traditional random-effect model. Furthermore, the proposed model does not separate the part-worth into aggregate time dependent and static individual level components as in Liechty et al. (2005) so that we do not need to make the restrictive assumption of an identical time pattern for all individuals. In addition, our model use demographic/firmographic variables to explain the variation on firm heterogeneity and time dependent part-worths, which cannot be done in a traditional dynamic linear model. We also allow a time varying impact matrix, which is used to uncover the changing characteristics of these impacts.

This study should have important implications for researchers and investors. Several studies in the literature have shown that investments in firms with high customer satisfaction lead to excess returns (e.g., Fornell et al. 2006; Aksoy et al. 2008). An argument commonly used to explain the phenomenon is that customer satisfaction is an important intangible asset which can influence shareholder value. Thus based on these studies, people may draw the conclusion that increases in the level of customer satisfaction always lead to increases in stock returns or shareholder values. Yet, as demonstrated in this paper, shareholder value is actually determined not only by the level of customer satisfaction but also by the association between customer satisfaction and shareholder value. An understanding of the dynamic firm-level association proposed by this paper will help researchers and investors further investigate the excess returns phenomenon as well as make better investment decisions over time.

Our study also provides important managerial implications for practitioners. First of all, the varying association suggests that shareholders do not fix their valuations on firms' customer satisfaction strategies. Therefore, firms need to adjust their customer satisfaction strategy not only to meet their customers' changing preference but also to impress their shareholders. This implies that firms have to consistently invest in customer satisfaction because an occasional high customer satisfaction does not have a lasting effect on shareholders' valuations. The dynamic association indicates that shareholders have the opportunity to adjust their valuations and appreciate firms' efforts and strategic flexibility in building customer satisfaction. Second, our result shows that there exists an insignificant association between customer satisfaction and shareholder value for some firms in any given year. Firms in those cases should investigate the possible causes and probably shift more resources to influence their shareholders' valuations.

In this study, we use the proposed model to address a marketing phenomenon. However, the model can be applied to address similar problems in other disciplines. When variable selection is the focus of a study, the model can be modified by incorporating indicator variables in the prior specification within the hierarchical structure to formally perform variable selection. Indeed, more work can be done to generalize the model for various applications.

Appendix

Appendix A. Derivation of Full Conditional Distributions

1. For $t = 1$, the prior distributions are:

$$\pi\left(\sigma_1^{-2}\right) \propto \left(\sigma_1^{-2}\right)^{p-1} \exp\left\{-\frac{\sigma_1^{-2}}{q}\right\},$$
$$\pi\left(C_1^{-1}\right) \propto |C_1^{-1}|^{\frac{b-J-1}{2}} \exp\left\{-\frac{1}{2}tr\left(\frac{1}{a}C_1^{-1}\right)\right\},$$
$$\pi\left(\text{vec}\left(\Delta_1\right)\right) \propto \exp\left\{-\frac{1}{2}(\text{vec}\left(\Delta_1\right) - \eta_0)'\gamma^{-1}(\text{vec}\left(\Delta_1 - \eta_0\right))\right\}.$$

1.1 For $i = 1, \ldots, N$, the full conditional distribution of β_{i1} is:

$p(\beta_{i1}|$ all others)

$$\propto \exp\left\{-\frac{1}{2}\left[\sigma_1^{-2}\left(Y_{i1}-X_{i1}^{'}\beta_{i1}\right)^2+\left(\beta_{i1}-\Delta_1 Z_{i1}\right)^{'}\left(\sigma_1^2 C_1\right)^{-1}\left(\beta_{i1}-\Delta_1 Z_{i1}\right)\right]\right\}$$
$$\propto \exp\left\{-\frac{\sigma_1^{-2}}{2}\left[\beta_{i1}^{'}\left(X_{i1}X_{i1}^{'}+C_1^{-1}\right)\beta_{i1}-2\beta_{i1}^{'}\left(X_{i1}Y_{i1}+C_1^{-1}\Delta_1 Z_{i1}\right)\right]\right\}$$

This expression is proportional to a normal density $N(\beta_{i1}^0, \Psi_{i1})$ where:

$\Psi_{i1} = \sigma_1^2(X_{i1}X_{i1}^{'}+C_1^{-1})^{-1}$ and $\beta_{i1}^0 = (X_{i1}X_{i1}^{'}+C_1^{-1})^{-1}(Y_{i1}X_{i1}+C_1^{-1}\Delta_1 Z_{i1})$.

1.2 Let $\eta_1 = \text{vec}(\Delta_1)$ and $Z_{i1}^* = I_J \otimes Z_{i1}^{'}$, where $\otimes$ represents the Kronecker product, then the full conditional distribution of η_1 is:

$$p(\eta_1|\text{ all others})$$
$$\propto \exp\left\{-\frac{1}{2}\left[(\eta_1-\eta_0)^{'}\gamma^{-1}(\eta_1-\eta_0)+\sum_{i=1}^{N}\left(\left(\beta_{i1}-Z_{i1}^*\eta_1\right)^{'}\left(\sigma_1^2 C_1\right)^{-1}\left(\beta_{i1}-Z_{i1}^*\eta_1\right)\right)\right]\right\}$$
$$\propto \exp\left\{-\frac{1}{2}\eta_1^{'}\left(\gamma^{-1}+\sigma_1^{-2}\sum_{i=1}^{N}\left(Z_{i1}^{*'}C_1^{-1}Z_{i1}^*\right)\right)\eta_1 - \eta_1^{'}\left(\gamma^{-1}\eta_0+\sigma_1^{-2}\sum_{i=1}^{N}(Z_{i1}^{*'}C_1^{-1}\beta_{i1})\right)\right\}.$$

This expression is proportional to a normal density $N(\eta_1^0, \Theta_1)$ where:

$$\Theta_1 = \left(\gamma^{-1}+\sigma_1^{-2}\sum_{i=1}^{N}\left(Z_{i1}^{*'}C_1^{-1}Z_{i1}^*\right)\right)^{-1},$$

$$\eta_1^0 = \Theta_1\left(\gamma^{-1}\eta_0+\sigma_1^{-2}\sum_{i=1}^{N}(Z_{i1}^{*'}C_1^{-1}\beta_{i1})\right).$$

1.3 The full conditional distribution of σ_1^{-2} is:

$$p(\sigma_1^{-2}|\text{ all others}) \propto \left(\sigma_1^{-2}\right)^{\frac{N(J+1)}{2}+p-1}$$

$$\times \exp\left\{-\sigma_1^{-2}\left[\frac{1}{q}+\frac{1}{2}\sum_{i=1}^{N}\left(\left(Y_{i1}-X_{i1}'\beta_{i1}\right)^2+\left(\beta_{i1}-\Delta_1 Z_{i1}\right)' C_1^{-1}\left(\beta_{i1}-\Delta_1 Z_{i1}\right)\right)\right]\right\}.$$

This expression is proportional to a Gamma density $\mathrm{Ga}(\frac{1}{2}N(J+1)+p, Q_1^{-1})$ where:

$$Q_1 = \frac{1}{2}\sum_{i=1}^{N}\Big((Y_{i1}-X_{i1}'\beta_{i1})^2+(\beta_{i1}-\Delta_1 Z_{i1})' C_1^{-1}(\beta_{i1}-\Delta_1 Z_{i1})\Big)+\frac{1}{q}.$$

1.4 The full conditional distribution of C_1^{-1} is:

$$\begin{aligned}
&p(C_1^{-1}|\text{ all others})\\
\propto&|C_1^{-1}|^{\frac{N+b-J-1}{2}}\exp\left\{-\frac{1}{2}tr(\frac{1}{a}C_1^{-1})-\frac{\sigma_1^{-2}}{2}\sum_{i=1}^{N}\Big((\beta_{i1}-\Delta_1 Z_{i1})' C_1^{-1}(\beta_{i1}-\Delta_1 Z_{i1})\Big)\right\}\\
\propto&|C_1^{-1}|^{\frac{N+b-J-1}{2}}\exp\left\{-\frac{1}{2}tr\left[C_1^{-1}\left(\frac{1}{a}I_J+\sigma_1^{-2}\sum_{i=1}^{N}\Big((\beta_{i1}-\Delta_1 Z_{i1})(\beta_{i1}-\Delta_1 Z_{i1})'\Big)\right)\right]\right\}.
\end{aligned}$$

This expression is proportional to a Wishart density $W(N+b,\ P_1^{-1})$ where $P_1=\frac{1}{a}I_J+\sigma_1^{-2}\sum_{i=1}^{N}(\beta_{i1}-\Delta_1 Z_{i1})(\beta_{i1}-\Delta_1 Z_{i1})'$.

1.5 Next, we derive the posterior distribution of $(\sigma_1^{-2}, C_1^{-1}, \Delta_1)$. If we substitute Equation (2.13) into Equation (2.12), we will obtain

$$Y_{it}=X_{it}'\Delta_t Z_{it}+X_{it}'\delta_{it}+\varepsilon_{it}=X_{it}'\Delta_t Z_{it}+\zeta_{it},$$

where $\zeta_{it}\sim N\Big(0,\sigma_t^2(X_{it}'C_tX_{it}+1)\Big)$. So the required posterior distribution is:

$$\begin{aligned}
&p_1(\sigma_1^{-2}, C_1^{-1}, \Delta_1|Y_1, X_1, Z_1)\\
\propto&p(Y_1|X_1, Z_1, \sigma_1^{-2}, C_1^{-1}, \Delta_1)\pi\left(\sigma_1^{-2}\right)\pi\left(C_1^{-1}\right)\pi\left(\Delta_1\right)
\end{aligned}$$

$$\propto \left(\sigma_1^{-2}\right)^{\frac{N}{2}+p-1} |C_1^{-1}|^{\frac{b-J-1}{2}} \prod_{i=1}^{N}(1+X_{i1}^{'}C_1X_{i1})^{-\frac{1}{2}}$$
$$\times \exp\left\{-\frac{\sigma_1^{-2}}{q}-\frac{1}{2}\left[tr(\frac{1}{a}C_1^{-1})+(\text{vec}(\Delta_1)-\eta_0)^{'}\gamma^{-1}(\text{vec}(\Delta_1)-\eta_0)\right.\right.$$
$$\left.\left.+\sigma_1^{-2}\sum_{i=1}^{N}\frac{(Y_{i1}-X_{i1}^{'}Z_{i1}^{*}\text{vec}(\Delta_1))^2}{1+X_{i1}^{'}C_1X_{i1}}\right]\right\}.$$

2. For $t=2$, we use the posterior distribution of the parameters in Year 1, which is shown above, as the prior distribution of the parameters in Year 2, i.e. $p_1(\sigma_2^{-2}, C_2^{-1}, \Delta_2|Y_1, X_1, Z_1)$. We now derive the full conditional distributions for Year 2:

   2.1 Similar to 1.1, $p(\beta_{i2}|$ all others) is a normal density $N(\beta_{i2}^0, \Psi_{i2})$ where:

   $$\Psi_{i2}=\sigma_2^2(X_{i2}X_{i2}^{'}+C_2^{-1})^{-1} \text{ and } \beta_{i2}^0=(X_{i2}X_{i2}^{'}+C_2^{-1})^{-1}(X_{i2}Y_{i2}+C_2^{-1}\Delta_2Z_{i2}).$$

   2.2 Let $\eta_2=\text{vec}(\Delta_2)$ and $Z_{i2}^{*}=I_J\otimes Z_{i2}^{'}$. The full conditional distribution of η_2 is:

   $$p(\eta_2|\text{ all others})\propto \exp\left\{-\frac{1}{2}\left[(\eta_2-\eta_0)^{'}\gamma^{-1}(\eta_2-\eta_0)\right.\right.$$
   $$\left.\left.+\sum_{i=1}^{N}\left(\sigma_2^{-2}\frac{(Y_{i1}-X_{i1}^{'}Z_{i1}^{*}\eta_2)^2}{1+X_{i1}^{'}C_2X_{i1}}+(\beta_{i2}-Z_{i2}^{*}\eta_2)^{'}(\sigma_2^2C_2)^{-1}(\beta_{i2}-Z_{i2}^{*}\eta_2)\right)\right]\right\}.$$

   This expression is proportional to a normal density $N_{JK}(\eta_2^0, \Theta_2)$ where:

   $$\Theta_2=\left(\gamma^{-1}+\sigma_2^{-2}\sum_{i=1}^{N}\left[Z_{i2}^{*'}C_2^{-1}Z_{i2}^{*}+\frac{Z_{i1}^{*'}X_{i1}X_{i1}^{'}Z_{i1}^{*}}{1+X_{i1}^{'}C_2X_{i1}}\right]\right)^{-1},$$

   $$\eta_2^0=\Theta_2\left[\gamma^{-1}\eta_0+\sigma_2^{-2}\sum_{i=1}^{N}\left(Z_{i2}^{*'}C_2^{-1}\beta_{i2}+\frac{Z_{i1}^{*'}X_{i1}Y_{i1}}{1+X_{i1}^{'}C_2X_{i1}}\right)\right].$$

2.3 The full conditional distribution of σ_2^{-2} is:

$$
\begin{aligned}
&p(\sigma_2^{-2}| \text{ all others}) \\
&\propto \left(\sigma_2^{-2}\right)^{\frac{N}{2}+p-1} \exp\left\{-\frac{\sigma_2^{-2}}{q} - \frac{1}{2}\sigma_2^{-2}\sum_{i=1}^{N}\frac{(Y_{i1} - X_{i1}'Z_{i1}^{*}\text{vec}(\Delta_2))^2}{1 + X_{i1}'C_2X_{i1}}\right\}\left(\sigma_2^{-2}\right)^{\frac{N}{2}}\left(\sigma_2^{-2}\right)^{\frac{NJ}{2}} \\
&\times \exp\left\{-\frac{1}{2}\sum_{i=1}^{N}\left[\sigma_2^{-2}\left(Y_{i2} - X_{i2}'\beta_{i2}\right)^2 + \left(\beta_{i2} - \Delta_2 Z_{i2}\right)'\left(\sigma_2^2 C_2\right)^{-1}\left(\beta_{i2} - \Delta_2 Z_{i2}\right)\right]\right\}.
\end{aligned}
$$

It is proportional to a Gamma density Ga $\left(\frac{1}{2}N(J+2)+p, Q_2^{-1}\right)$ where:

$$
\begin{aligned}
Q_2 =& \frac{1}{q} + \frac{1}{2}\sum_{i=1}^{N}\left(\frac{(Y_{i1} - X_{i1}'Z_{i1}^{*}\text{vec}(\Delta_2))^2}{1 + X_{i1}'C_2X_{i1}} + \left(Y_{i2} - X_{i2}'\beta_{i2}\right)^2\right. \\
&\left. + \left(\beta_{i2} - \Delta_2 Z_{i2}\right)' C_2^{-1}\left(\beta_{i2} - \Delta_2 Z_{i2}\right)\right)
\end{aligned}
$$

2.4 The full conditional distribution of C_2^{-1} is:

$$
\begin{aligned}
&p(C_2^{-1}| \text{ all others}) \\
\propto& |C_2^{-1}|^{\frac{N+b-J-1}{2}}\left[\prod_{i=1}^{N}(1 + X_{i1}'C_2X_{i1})^{-\frac{1}{2}}\right]\exp\left[-\frac{1}{2a}tr(C_2^{-1})\right] \\
&\times \exp\left[-\frac{\sigma_2^{-2}}{2}\sum_{i=1}^{N}\left((\beta_{i2} - \Delta_2 Z_{i2})'C_2^{-1}(\beta_{i2} - \Delta_2 Z_{i2}) + \frac{(Y_{i1} - X_{i1}'\Delta_2 Z_{i1})^2}{1 + X_{i1}'C_2X_{i1}}\right)\right]
\end{aligned}
$$

Since the full conditional distribution of C_2^{-1} is not a standard probability density, we will generate random deviates from this distribution using the Metropolis-Hastings algorithm with the following Wishart proposal density,

$$
W\left(N+b, \left(\frac{1}{a}I_J + \sigma_2^{-2}\sum_{i=1}^{N}\left[(\beta_{i2} - \Delta_2 Z_{i2})(\beta_{i2} - \Delta_2 Z_{i2})'\right]\right)^{-1}\right).
$$

An acceptance probability (Chib & Greenberg, 1995) is used to determine if the generated matrix will be accepted.

2.5 The posterior distribution of $(\sigma_2^{-2}, C_2^{-1}, \Delta_2|D_2)$ is:

$$
\begin{aligned}
& p_2(\sigma_2^{-2}, C_2^{-1}, \Delta_2|D_2) \\
\propto & p(Y_2|X_2, Z_2, \sigma_2^{-2}, C_2^{-1}, \Delta_2) p_1(\sigma_2^{-2}, C_2^{-1}, \Delta_2|D_1) \\
\propto & \left(\sigma_2^{-2}\right)^{\frac{2N}{2}+p-1} |C_2^{-1}|^{\frac{b-J-1}{2}} \prod_{i=1}^{N}\prod_{s=1}^{2}\left(1 + X_{is}^{'} C_2 X_{is}\right)^{-\frac{1}{2}} \\
& \times \exp\Bigg\{ -\frac{\sigma_2^{-2}}{q} - \frac{1}{2}\Bigg[tr(\frac{1}{a}C_2^{-1}) + (\text{vec}(\Delta_2 - \eta_0))' \gamma^{-1} (\text{vec}(\Delta_2 - \eta_0)) \\
& + \sigma_2^{-2} \sum_{i=1}^{N}\sum_{s=1}^{2} \frac{(Y_{is} - X_{is}' Z_{is}^{*} \text{vec}(\Delta_2))^2}{1 + X_{is}' C_2 X_{is}} \Bigg]\Bigg\}
\end{aligned}
$$

3. Similar arguments then lead to the following full conditional distributions for $t > 2$:

3.1 For $i = 1, \ldots, N$, $p(\beta_{it}|$ all others$) = N_J(\beta_{it}^0, \Psi_{it})$, where:

$$
\begin{aligned}
\Psi_{it} =& \sigma_t^2 (X_{it} X_{it}' + C_t^{-1})^{-1}, \\
\beta_{it}^0 =& (X_{it} X_{it}' + C_t^{-1})^{-1} (Y_{it} X_{it} + C_t^{-1} \Delta_t Z_{it});
\end{aligned}
$$

3.2 $p(\text{vec}(\Delta_t)|$ all others$) = N_{JK}(\eta_t^0, \sigma_t^2 \Theta_t^{-1})$, where:

$$
\begin{aligned}
\Theta_t =& \left[\gamma^{-1} + \sigma_t^{-2} \sum_{i=1}^{N} \left(\bar{Z}_{it}' C_t^{-1} \bar{Z}_{it} + \sum_{s=1}^{t-1} \frac{(X_{is} X_{is}' \otimes Z_{is} Z_{is}')}{X_{is}' C_t X_{is} + 1} \right)\right]^{-1}, \\
\eta_t^0 =& \Theta_t^{-1} \left[\gamma^{-1} \eta_0 + \sigma_t^{-2} \sum_{i=1}^{N} \left(\bar{Z}_{it}' C_t^{-1} \beta_{it} + \sum_{s=1}^{t-1} \frac{Y_{is} X_{is} \otimes Z_{is}}{X_{is}' C_t X_{is} + 1} \right)\right], \\
\bar{Z}_{it} =& I_J \otimes Z_{it}';
\end{aligned}
$$

3.3 $p(\sigma_t^{-2}|\text{ all others}) = \text{Ga}(\frac{1}{2}N(J+t)+p, Q_t^{-1})$, where:

$$Q_t = \frac{1}{q} + \frac{1}{2}\sum_{i=1}^{N}\Bigg[(Y_{it} - X_{it}'\beta_{it})^2 + (\beta_{it} - \Delta_t Z_{it})' C_t^{-1}(\beta_{it} - \Delta_t Z_{it}) + \sum_{s=1}^{t-1}\frac{(Y_{is} - X_{is}'\Delta_t Z_{is})^2}{X_{is}'C_t X_{is} + 1}\Bigg];$$

3.4 $p(C_t^{-1}|\text{ all others})$

$$\propto |C_t|^{-\frac{N+b-J-1}{2}}\left[\prod_{i=1}^{N}\prod_{s=1}^{t-1}(X_{is}'C_t X_{is}+1)^{-\frac{1}{2}}\right]\exp\left\{-\frac{1}{2}tr(a^{-1}C_t^{-1})\right\}$$
$$\times \exp\Bigg\{-\frac{1}{2\sigma_t^2}\sum_{i=1}^{N}\Bigg[(\beta_{it} - \Delta_t Z_{it})' C_t^{-1}(\beta_{it} - \Delta_t Z_{it}) + \sum_{s=1}^{t-1}\frac{(Y_{is} - X_{is}'\Delta_t Z_{is})^2}{X_{is}'C_t X_{is} + 1}\Bigg]\Bigg\},$$

which is not a standard probability density. Random deviates from the distribution can be generated using the Metropolis-Hastings algorithm with the following Wishart proposal density,

$$W\left(N+b, \left(\frac{1}{a}I_J + \sigma_t^{-2}\sum_{i=1}^{N}(\beta_{it} - \Delta_t Z_{it})(\beta_{it} - \Delta_t Z_{it})'\right)^{-1}\right).$$

An acceptance probability is used to determine if the generated matrix will be accepted.

A random sample from the joint posterior distribution can be obtained by generating random deviates iteratively and recursively according to the above full conditional distributions. We have simulated $25,000$ iterations, out of which the last $20,000$ iterations are used for generating parameter estimates. Convergence was checked by starting the chain from multiple initial values and by the inspection of trace plots. In addition to the plot assessment, we have implemented a standard CODA function that provides some convergence diagnostics (Cowles, 1996; Newton & Raftery, 1994). The Gelman and Rubin (1992) diagnostic is also employed which compares the between-chain and within-chain dispersions, akin to an analysis of variance. See Cowles and Carlin (1996) for a review of convergence statistics.

Appendix B. Calculation of the Log Marginal Likelihood

To calculate the Bayes factor comparing one model with another model, one needs to calculate the (log) marginal likelihoods. We apply two methods in the calculation. The first method is the harmonic mean, which is easy to compute:

$$\text{Harmonic Mean: } \hat{p}(Y|M_k) = \left(\sum_{i=1}^{m} \frac{1}{p(Y|\theta^i, M_k)}\right)^{-1},$$

where $\hat{p}(Y|M_k)$ denote an estimate of the marginal likelihood of HBDM or MJB and $p(Y|\theta^i, M_k)$ is the likelihood function of HBDM or MJB based on parameter estimates sampled from its posterior distribution in the ith iteration.

The second one is the fourth method listed in Newton and Raftery (1994) which has been used by various marketing researchers. Let δ be a small number and m be the total number of iterations in MCMC. Then, the marginal likelihood can be estimated by iteratively solving the following equation:

$$\text{Newton and Raftery: } \hat{p}(Y|M_k) = \frac{\frac{\delta m}{1-\delta} + \sum_{i=1}^{m} \frac{p(Y|\theta^i, M_k)}{\delta \hat{(p)}(Y|M_k) + (1-\delta) p(Y|\theta^i, M_k)}}{\frac{\delta m}{(1-\delta)\hat{(p)}(Y|M_k)} + \sum_{i=1}^{m} \frac{1}{\delta \hat{(p)}(Y|M_k) + (1-\delta) p(Y|\theta^i, M_k)}}.$$

For DIC, we first define the deviance of model M_k as:

$$D(\Theta|M_k) = -2log(p(Y|\Theta, M_k)) + 2log(f(y)),$$

Then, the expectation $\bar{D}(\Theta|M_k) = E^{\Theta}(D(\Theta|M_k))$ evaluates how well the model fit the data. The effective number of parameters of the model is then computed as $p_D(M_k) = \bar{D}(\Theta|M_k) - D(Th\bar{e}ta|M_k)$, where $\bar{\Theta}$ is the expectation of Θ. The value of DIC will be a sum of the effective number of parameters and the goodness-of-fit of the model. Mathematically,

$$DIC_{M_k} = p_D(M_k) + \bar{D}(\Theta|M_k).$$

Chapter 3

A Bayesian Random Coefficient Multinomial Probit Model for the Analysis of Consumer Choice Panel Data

3.1 Introduction

Rossi, McCulloch, and Allenby (1996) (RMA henceforth) proposed a random coefficient multinomial probit model to evaluate the value of purchase history data in target marketing. Their results indicate that there exists potential for improving the profitability of direct marketing efforts by more fully utilizing household purchase history for micro-marketing. There has since been a growing literature to investigate the mechanism and benefits of one-to-one marketing: for example, store level customization (Montgomery, 1997), the necessity of customer targeting (Chen, Narasimhan, & Zhang, 2001), e-customization (Ansari & Mela, 2003), the timing of customized promotion in online stores (J. Zhang & Krishnamurthi, 2004), customer satisfaction (Bodapati & Gupta, 2003), one-to-one marketing in B-2-B environment (Ghosh, Dutta, & Stremersch, 2006), the ideal level of customization (J. Zhang & Wedel, 2009), etc. (See also Rossi, Allenby, & McCulloch, 2006).

To analyze a consumer choice panel data set, RMA include household demographic information in their analysis and conclude that the demographic characteristics barely explain the variability in individual level parameter estimates. However, when we apply the Bayesian model in RMA to reanalyze the tuna purchase data in that paper, we find that some demographic variables have explanatory power and that their Bayesian estimates are unstable when the prior variances are set larger and larger. Whereas subjective information may be used to specify informative priors, in many cases there is no strong prior belief so that situations to "let the data speak" are preferred. In such cases, *vague priors* (e.g., proper priors with large variances) are commonly employed, and a model which is not sensitive to specific vague prior specifications is desirable (for related work, see Hobert & Casella, 1996; Berger, 2000; Sun et al., 2001; Bayarri & Berger, 2004; and Berger et al., 2005). Thus, we build upon the model in RMA and propose a new Bayesian random coefficient multinomial probit model to analyze consumer choice panel data. The proposed model has several attractive features. First, it is more general than the model in RMA (delineated as the benchmark model henceforth) and so it is applicable even when the RMA model is not. Second, our proposed model appears to be less sensitive to prior specifications with large variances, and it always yields a proper posterior distribution. Third, the algorithm used to implement the proposed model is approximately a Gibbs sampling, and it is very efficient in computing the required Bayesian estimates.

This chapter is organized as follows. Section 3.2 summarizes a class of popular models and algorithms used to analyze choice data. We then present our proposed

Bayesian random coefficient multinomial probit model together with a description of the computational algorithm utilized to implement the procedure in Section 3.3. In Section 3.4, we perform a simulation study to compare the performance of the proposed model and the RMA benchmark model under different scenarios. Section 3.5 reanalyzes the RMA tuna purchase data and compares our results with those using the benchmark model. In Section 3.6, we summarize the contribution of work and provide a conclusion.

3.2 Literature Review

3.2.1 Bayesian Multinomial Probit Model (MNP)

To avoid the well known IIA problem associated with a multinomial logit model, a number of scholars have proposed the multinomial probit (MNP) model to analyze discrete choice data (McCulloch & Rossi, 1994; Nobile, 1998; McCulloch, Polson, & Rossi, 2000; Imai & Dyk, 2005). An excellent review of the MNP model for marketing applications is given in Zeithammer & Lenk (2006).

Suppose that there are m brands and H consumers. Let U_i be the observed choice for consumer i such that $U_i \in \{0, 1, 2, \ldots, m-1\}$, $i = 1, \ldots, H$, where one of the brands is labeled as brand 0. We call it *the base brand* in this chapter. In an MNP model, choice U_i is modeled in terms of a latent vector W_i with length $m-1$ via the following equation:

$$U_i(W_i) = \begin{cases} 0, & \text{if} \max(W_i) < 0, \\ j, & \text{if} \max(W_i) = W_{ij} > 0, j = 1, \ldots, m-1. \end{cases} \tag{3.1}$$

where $\max(W_i)$ stands for the maximum element of the vector W_i, and W_{ij} represents the j^{th} component of W_i. Because of the existence of the base brand, the latent variable W_{ij} is measuring *the relative utility* between brand j and the base brand for the i^{th} consumer. Suppose that there is a $m \times 1$ vector Y_i, the s^{th} component of which measures *the absolute utility* of brand s for consumer i. If we select brand m as the base brand, then $W_{ij} = Y_{ij} - Y_{im}$.

For the i^{th} consumer, let X_i be a $m \times k$ matrix containing values of k covariates for all m brands and R_i be a $(m-1) \times k$ matrix defined by $R_{ij} = X_{ij} - X_{im}$, where X_{ij} is the j^{th} row of X_i. It is generally assumed in MNP that:

$$W_i = R_i\beta + \epsilon_i, \tag{3.2}$$

where β is a $k \times 1$ vector of coefficients, and $\epsilon_i \sim N(0, \Sigma)$, where $\Sigma = (\sigma_{st})$ is a positive definite matrix of dimension $(m-1) \times (m-1)$.

It is well known that the above model is not identifiable and the identification constraint $\sigma_{11} = 1$ is typically applied. However, such constraint complicates the Bayesian analysis as one has to deal with the restricted covariance matrix which is not easy to generate in an MCMC algorithm to compute the required parameter estimates. Researchers have developed various methods to solve this problem. McCulloch and Rossi (1994) propose an algorithm to work with non-identified parameters and then perform post-processing to obtain parameter estimates. Let (β, Σ) represent the non-identified parameters. The model employs a normal prior on β, $\beta \sim N(b_0, A^{-1})$ and a Wishart prior on Σ^{-1}, $\Sigma^{-1} \sim W(\nu, V)$, where $E(\Sigma^{-1}) = \nu V$. Since both are conjugate priors, the full conditional distributions are all standard probability distributions and can utilize a Gibbs sampling algorithm to generate a random sample from the posterior distribution. Although β and Σ are non-identified parameters, $\beta/\sqrt{\sigma_{11}}$ and Σ/σ_{11} are identified. So, posterior estimates of the identified parameters,

$(\beta/\sqrt{\sigma_{11}},\ \Sigma/\sigma_{11})$, can be computed after obtaining draws of β and Σ^{-1} from the Gibbs sampler. Nobile (1998) modifies McCulloch and Rossi's model by introducing a "hybrid Markov chain" so that the computational performance of this model is improved.

McCulloch, Polson and Rossi (2000) present a method to deal with fully identified parameters. First, they reparameterize Σ and the error term in Equation (3.2) as follows:

$$\Sigma = \begin{pmatrix} 1 & \gamma' \\ \gamma & \Phi + \gamma\gamma' \end{pmatrix}, \quad \epsilon_i = \begin{pmatrix} u_i \\ z_i \end{pmatrix}, \tag{3.3}$$

where u_i is a scalar and z_i is a $(m-2) \times 1$ vector, the authors show that:

$$u_i \sim N(0, \sigma_{11}) = N(0, 1), \tag{3.4}$$

$$z_i|u_i \sim N(\gamma/\sigma_{11} u_i, \Phi) = N(\gamma u_i, \Phi). \tag{3.5}$$

Then, they assume the conjugate prior distributions on γ and Φ^{-1}, which are:

$$\gamma \sim N(\bar{\gamma}, B^{-1}), \tag{3.6}$$

$$\Phi^{-1} \sim W(\kappa, C). \tag{3.7}$$

Under these priors, the full conditional distributions of γ and Φ^{-1} can be derived, which are standard probability distributions, namely, a normal distribution and a Wishart distribution, respectively. However, the authors mention that their Gibbs sampling algorithm can converge slowly and that the sampler may get stuck when vague improper priors are used (p. 183 of their paper).

Imai and van Dyk (2005) propose a method which works on the fully identified parameters using marginal data augmentation. In addition to the identified parameter set (W_i, β, Σ) (i.e., $\sigma_{11} = 1$), they consider an augmented parameter set $(\tilde{W}_i, \tilde{\beta}, \tilde{\Sigma}, \alpha)$, where:

$$\tilde{W}_i = \alpha W_i, \tag{3.8}$$

$$\tilde{\beta} = \alpha \beta, \tag{3.9}$$

$$\tilde{\Sigma} = \alpha^2 \Sigma, \tag{3.10}$$

and α is a scalar. Note that the augmented parameter $\tilde{\Sigma}$ is an unconstrained positive definite matrix, and α^2 is the $(1,1)$ element of $\tilde{\Sigma}$, denoted as $\tilde{\sigma}_{11}$. Beginning with a conjugate prior on $\tilde{\Sigma}^{-1} \sim W(\nu, \tilde{S})$, the authors prove that the conditional distribution of $\alpha^2|\Sigma \sim \alpha_0^2 \text{tr}(\Sigma^{-1}S^{-1})/\chi^2_{\nu(m-1)}$, where α_0 is a positive constant. This yields the marginal prior distribution of Σ as shown below:

$$p(\Sigma) \propto |\Sigma|^{-(\nu+m)/2} \left[\text{tr}(\Sigma^{-1}S^{-1})\right]^{-\nu(m-1)/2}, \tag{3.11}$$

where $S = \tilde{S}/\alpha_0^2$. This is a proper distribution when $\nu > m - 1$. The paper then proposes two sampling schemes to implement this method. The authors show that their algorithms run as fast as Nobile's (1998) algorithm, and are faster than the algorithms of McCulloch and Rossi (1994) and McCulloch et al. (2000). It has an advantage over McCulloch and Rossi (1994) as it is not sensitive to starting values. Moreover, while Nobile (1998) needs to add an additional Metropolis step to improve the computational efficiency, Imai and van Dyk (2005) do not require a Metropolis step; therefore, their method is easier to implement.

All these MNP models deal with cross-sectional choice data only and it is not clear how they can be extended to analyze panel data on consumer choices when heterogeneity is present. Although McCulloch and Rossi (1994) discuss how to generalize their model to analyze panel data, they assume a common normal prior for the individual level β_i's which is not helpful in understanding or explaining heterogeneity.

3.2.2 Random Coefficient Multinomial Probit Model (RC-MNP)

Rossi, McCulloch, and Allenby (1996) develop a RC-MNP model to analyze a panel data set on household purchasing cans of tuna. Let I_{ht} denote the brand choice for household h at time t, where $h = 1, \ldots, H$ and $t = 1, \ldots, T_h$. Household h will choose brand j, $j = 1, \ldots, m$, if the corresponding brand utility is the largest amongst the m brands. More specifically, let y_{ht} be an m-dimensional vector representing the latent utilities; then:

$$I_{ht}(y_{ht}) = j, \quad \text{if } y_{ht}[j] > max(y_{ht}[-j]), \tag{3.12}$$

where $y_{ht}[j]$ is the j^{th} component of y_{ht} and $max(y_{ht}[-j])$ stands for the maximum value of all components in y_{ht} excluding $y_{ht}[j]$. The latent utility, y_{ht}, is assumed to follow a multivariate regression:

$$y_{ht} = X_{ht}\beta_h + \varepsilon_{ht}, \quad \varepsilon_{ht} \sim N(0, \Lambda), \tag{3.13}$$

where X_{ht} is an $m \times k$ matrix of choice characteristics. β_h is a k-dimensional vector containing household h's sensitivity to the marketing mix variables, and it also includes a brand-specific intercept term for each of the m brands. ε_{ht} is an error term that follows a multivariate normal distribution with mean 0 and covariance matrix Λ, which is a diagonal matrix with diagonal elements $(\sigma_1^2, \ldots, \sigma_m^2)$. To solve identification problems associated with the RC-MNP model, RMA fix the intercept parameter of the first brand at 0 and set the first element of the covariance matrix, σ_1^2 to 1.

In addition, the authors assume a multivariate regression model for β_h using demographic variables Z_h as independent variables:

$$\beta_h = \Delta Z_h + \delta_h, \quad \delta_h \sim N(0, V_\beta), \tag{3.14}$$

where the matrix Δ measures the impact of demographic variables on households' sensitivity to brand features.

The use of a diagonal covariance matrix in Equation (3.13) is quite restrictive. First of all, it basically assumes that the brand utilities are independent from each other; however, it is more common to have correlated utilities, even they are latent. Second, according to the model setting, a brand is selected when the associated utility is the largest, so the selection is actually based on *the relative utilities.* Without loss of generality, we can assume that the brand with the largest variance is the first brand (and $\sigma_1^2 = 1$) so that $\sigma_j^2 < 1$ for all $j > 1$ (we can achieve this by suitably permuting the index of brands). If we set brand m as the base brand, and denote the relative utility between brand j and the base brand for household h at time t by $w_{ht}[j] = y_{ht}[j] - y_{ht}[m]$, where $j = 1, \ldots, m-1$, then the covariance structure of w_{ht} will be:

$$\begin{pmatrix} 1+\sigma_m^2 & \sigma_m^2 & \cdots & \sigma_m^2 \\ \sigma_m^2 & \sigma_2^2+\sigma_m^2 & \cdots & \sigma_m^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_m^2 & \sigma_m^2 & \cdots & \sigma_{m-1}^2+\sigma_m^2 \end{pmatrix}. \tag{3.15}$$

It is clear that the variance of $w_{ht}[j]$ is $\sigma_j^2 + \sigma_m^2$, which lies between 0 and 2, and the covariance of $w_{ht}[j]$ and $w_{ht}[k]$ is σ_m^2. Consequently, the covariance is always positive, and it remains the same for all pairs of relative utilities. Such positive equal covariance assumption for the RMA model appears to be overly restrictive, so it calls for a more general formulation of the random-coefficient multinomial probit model to relax the restrictions.

3.3 The Proposed Model

3.3.1 The Bayesian Random Coefficient Multinomial Probit Model

In this section, we propose a more general random coefficient multinomial probit model to analyze choice panel data. Let I_{ht} and y_{ht} represent the household choice and the corresponding latent utility as in Section 3.2.2. Similar to RMA, we assume that the brand with the highest utility will be selected, that is:

$$I_{ht}(y_{ht}) = j, \quad \text{if } y_{ht}[j] > max(y_{ht}[-j]), \tag{3.16}$$

We use a regression model to relate the latent utility with brand covariate X_{ht} as follows:

$$y_{ht} = X_{ht}\beta_h + \varepsilon_{ht}, \quad \varepsilon_{ht} \sim N(0, R), \tag{3.17}$$

where $R = (r_{ij})$ is a correlation matrix. As in RMA, we set the first element of β_h to be 0 to solve an identification problem. Note that we use the maximum allowable number of free variance-covariance parameters in our model such that it is still identifiable (see Albert & Chib, 1993). By this modification, our proposed model is more general than the RMA benchmark model as it allows the existence of non-zero correlations between brand utilities. Furthermore, the variance of the relative utility $w_{ht}[j]$ equals $2 + 2r_{jm}$, which lies between 0 and 4 since $-1 \leq r_{jm} \leq 1$. In addition, the covariance of $w_{ht}[j]$ and $w_{ht}[k]$ is $1 - r_{km} - r_{jm} + r_{jk}$, which allows for both positive and negative values, and it is clear that the covariances are not required to be identical.

Following RMA, we use demographic variables to model β_h in a multivariate regression setting. Since we set the first element of β_h to be 0, there are only $(k-1)$ effective parameters. For simplicity, we will use β_h to represent this reduced coefficient vector henceforth and $\tilde{X}_{ht}$ to represent the corresponding X matrix, which is obtained by deleting the first column of X_{ht}. Let Z_h be an l dimensional vector with one as the first element and the remaining $l-1$ elements being values of the demographic variables for household h. Our model for β_h is:

$$\beta_h = \Delta Z_h + \delta_h, \quad \delta_h \sim N(0, V_\beta), \tag{3.18}$$

where Δ measures the impact of demographic variables on the β coefficients. The error term, δ_h, follows a $(k-1)$-variate normal distribution with mean 0 and covariance matrix V_β.

3.3.2 Prior Specification

The following natural conjugate priors are assumed for Δ and V_β:

$$\Delta | V_\beta \sim \text{MN}(\Delta_0, V_\beta, A_d^{-1}); \;\; V_\beta^{-1} \sim W(v, V), \tag{3.19}$$

where $MN(\Delta_0, V_\beta, A_d^{-1})$ denotes a matrix normal distribution with mean Δ_0, row covariance matrix V_β, and column covariance matrix A_d^{-1} (Dawid, 1981), and $W(v, V)$ is a Wishart distribution with mean vV. For the correlation matrix R, we assume the locally uniform prior:

$$\pi(R) = \text{I}\{R \text{ is a positive definite correlation matrix}\}. \tag{3.20}$$

Since the space of $m \times m$ correlation matrices is convex and compact (Rousseeuw & Molenberghs, 1994), the prior distribution specified in Equation (3.20) is proper.

3.3.3 Full Conditional Distributions

For Bayesian computation, we generate an approximate sample from the joint posterior distribution by drawing random deviates iteratively and recursively from the following full conditional distributions. Their derivations are given in Appendix A. Some of these expressions are similar to those in RMA, but in comparing results we find that the full conditional distribution of V_β^{-1} in RMA is incorrect. See Appendix A, part 4 for details. Here,

$$y_{ht}| \text{ all others} \sim TN(\tilde{X}_{ht}\beta_h, R), \tag{3.21}$$

where $y_{ht}[I_{ht}] > \max(y_{ht}[-I_{ht}])$, and $\tilde{X}_{ht}$ is the reduced matrix of X_{ht} by deleting its first column. And,

$$\beta_h| \text{ all others} \sim N(V_\beta^n \Bigg[\sum_t (\tilde{X}'_{ht} R^{-1} y_{ht}) + V_\beta^{-1} \Delta Z_h \Bigg], V_\beta^n), \tag{3.22}$$

where $V_\beta^n = \Bigg[\sum_t (\tilde{X}'_{ht} R^{-1} \tilde{X}_{ht}) + V_\beta^{-1} \Bigg]^{-1}$. Also,

$$\Delta| \text{ all others} \sim MN((BZ' + \Delta_0 A_d)(ZZ' + A_d)^{-1}, V_\beta, (ZZ' + A_d)^{-1}), \tag{3.23}$$

where $B = \Bigg[\beta_1, \ldots, \beta_H \Bigg]$ and $Z = \Bigg[Z_1, \ldots, Z_N \Bigg]$. Finally,

$$V_\beta^{-1}| \text{ all others} \sim W(v+H+l, \Bigg[(B-\Delta Z)(B-\Delta Z)' + (\Delta-\Delta_0)A_d(\Delta-\Delta_0)' + V^{-1} \Bigg]^{-1}). \tag{3.24}$$

Note, the full conditional distribution for R is not a standard probability distribution. Sampling from this distribution is discussed in Section 3.3.4. We have also written a program in R to implement the proposed MCMC algorithm.

3.3.4 Sampling of Correlation Matrix

Unlike the covariance matrix, it is difficult to find a conjugate prior for R so that the corresponding full conditional distribution is a standard probability distribution. To draw random correlation matrix in Bayesian computation, some researchers have proposed sampling methods which are very computationally intensive. Chib and Greenburg (1998) assume that the off-diagonal components of R follow a multivariate normal distribution truncated to the space of correlation matrix, and then use the Metropolis-Hastings algorithm to sample new correlation matrices. The approach proposed by Liechty et al. (2004) also uses the Metropolis-Hastings algorithm, but they simulate elementwise each component of a new correlation matrix. Generally, the acceptance rates in these algorithms are low and so it may take a long time to run these procedures. Here, we use a parameter expansion technique (Liu & Daniels, 2006; Hobert & Marchev, 2008) to generate random correlation matrices. (Details of this simulation strategy are given in Appendix B.) Using this method, we can employ an inverse Wishart distribution to generate covariance matrix Σ which is transformed into R through a one-to-one mapping. The derived R is then accepted based on a Metropolis-Hastings acceptance probability. Typically, the acceptance rate of this algorithm is high and the acceptance probability is straightforward to compute, so the method is quite efficient.

3.4 Simulated Data Example

We perform a simulation study to compare the performance of our proposed model and the RMA benchmark model. Two sets of synthetic data are generated according to these models, respectively, namely, one is generated based on an assumption of diagonal covariance matrix for the brand utilities, and the other one using a correlation matrix assumption. We set $m = 5$ competing brands, $H = 100$ households, $k = 8$ with three marketing mix variables, and $l = 6$ with five demographic variables. We assume that each household has a history of $T = 10$ purchases. Then, we generate the impact matrix Δ from a matrix normal distribution and the inverse of the covariance matrix V_β from a Wishart distribution. For the data set based on the RMA benchmark model, we generate the reciprocal of each diagonal component of Λ (except the first one) from a Gamma distribution. For the data set based on our model, we set the correlation matrix R to be:

$$R = \begin{pmatrix} 1 & 0.2 & 0.1 & -0.1 & -0.2 \\ 0.2 & 1 & 0.2 & 0.1 & -0.1 \\ 0.1 & 0.2 & 1 & 0.2 & 0.1 \\ -0.1 & 0.1 & 0.2 & 1 & 0.2 \\ -0.2 & -0.1 & 0.1 & 0.2 & 1 \end{pmatrix}. \tag{3.25}$$

Next, β_h and brand utility y_{ht} are created from the corresponding multivariate normal distributions, and the consumer choice observations I_{ht} are determined according to the generated brand utilities.

To allow evaluation of prediction performance, we divide the data into two parts: a calibration set for parameter estimation and a validation set for purchase prediction. For each household, we assign the first 9 purchases to the calibration set and the last one to the validation set.

For both models, we assume two proper priors. For *diffuse proper priors I*, we set $A_d = 0.01E_6$, $v = 11$, $V = E_7$, $a = 3$, and $b = 1$, where E_l represents an $l \times l$ identity matrix, and a and b are the hyperparameters in the Gamma prior distribution of σ_j^2, with mean equals to ab. For *diffuse proper priors II*, we set $A_d = 0.001E_6$, $v = 11$, $V = E_7$, $a = 0.06$, and $b = 50$. Comparisons are done in terms of mean absolute deviation (MAD) and root mean squared error (RMSE) for parameter recovery. Results of parameter estimation accuracy between our model and the benchmark model are shown in Table 3.1. It is clear from the tables that our proposed model performs better than or equal to the benchmark model in all cases, even when the data are generated according to a diagonal covariance matrix. Please note that when diffuse proper priors II are applied, the benchmark model performs poorly yielding dramatically larger errors compared with those from using diffuse proper priors I. In contrast, our proposed model always gives similar errors in both cases indicating that the parameter estimates are stable under the two different prior settings.

Holdout sample prediction accuracy results under different priors for the two models are given in Table 3.2. The RMA benchmark model performs reasonably well when diffuse proper priors I are assumed, but the performance drops significantly when diffuse proper priors II are employed. However, our model again yields consistently better results and dominates the benchmark model in all cases.

3.5 Application: Tuna Scanner Data

In this section, we reanalyze the scanner panel dataset of tuna purchases considered in RMA (we thank Eric T. Bradlow and Greg M. Allenby for kindly providing the data). We compare the parameter estimation and model prediction results of our proposed

Table 3.1. Simulation Analysis - Parameter estimation accuracy comparison between the RMA benchmark model and our proposed model for calibration set under different data structures.

Data Structure	Prior Specification	Variable	Rossi et al. (1996) MAD	Rossi et al. (1996) RMSE	Proposed Model MAD	Proposed Model RMSE
Diagonal Covariance	Diffuse Proper Prior I	Y	75.54	107.95	**67.64**	**96.87**
		β	19.51	25.29	**17.36**	**22.47**
		$\mathrm{Cov}(\Lambda/R)$	3.46	6.72	**0.74**	**2.87**
		Δ	4.16	5.55	**3.74**	**4.94**
		V_β	0.09	0.12	0.09	**0.11**
	Diffuse Proper Prior II	Y	480.26	806.78	**67.55**	**96.72**
		β	187.54	259.12	**17.32**	**22.45**
		$\mathrm{Cov}(\Lambda/R)$	59464.52	74748.99	**0.74**	**2.87**
		Δ	35.25	44.18	**3.73**	**4.94**
		V_β	14479.28	19161.43	**0.09**	**0.12**
Correlation	Diffuse Proper Prior I	Y	75.44	107.96	**68.27**	**97.91**
		β	19.28	24.91	**17.35**	**22.39**
		$\mathrm{Cov}(\Lambda/R)$	0.24	0.32	**0.14**	**0.15**
		Δ	4.14	5.49	**3.75**	**4.95**
		V_β	0.09	0.12	0.09	**0.11**
	Diffuse Proper Prior II	Y	486.08	819.04	**67.12**	**96.29**
		β	191.71	268.39	**16.81**	**21.64**
		$\mathrm{Cov}(\Lambda/R)$	11375.68	35383.45	**0.14**	**0.15**
		Δ	35.49	44.95	**3.67**	**4.79**
		V_β	16183.05	21313.30	**0.09**	**0.12**

Note: **bold** indicates that the measurement (MAD or RMSE) of the proposed model is smaller than the benchmark model.

model vs. those from the RMA benchmark model under different prior specifications.

3.5.1 The Tuna Data Set

We start with a brief summary description of the tuna data in RMA (1996). The data set contains purchase information of some 400 households on five brands of tuna: Chicken of the Sea (Water), Starkist (Water), House Brand (Water), Chicken of the

Table 3.2. Simulation Analysis - Holdout sample prediction accuracy comparison between the RMA benchmark model and our proposed model.

Covariance Structure	Prior Specification	Rossi et al. (1996)	New Model
Diagonal Data	Diffuse Proper Prior I	100%	100%
	Diffuse Proper Prior II	78%	**100%**
Correlation Data	Diffuse Proper Prior I	99%	99%
	Diffuse Proper Prior II	72%	**99%**

Note: **bold** indicates that the prediction accuracy of the proposed model is better than the benchmark model.

Sea (Oil), and Starkist (Oil). Three purchasing characteristics are also available: the price, the existence of in-store displays, and feature advertisements at the moment of purchase. As in RMA, price is recorded in its logarithm form, and display and feature are coded as dummy variables. In addition, the data set also provides five household demographic characteristics: household income, family size, if head of household is retired, if head of household is employed, and if head of household is single mom. Further details of this dataset can be found in RMA (1996).

As with our simulation study, we divide the full data set into two parts: the calibration set and the validation set to evaluate the models' predictive validity. The last three purchases of each household are saved to form the validation set, and the earlier observations are used in the calibration set for parameter estimation.

3.5.2 Results and Findings

One of the goals in RMA is to investigate the relative value of demographic vs. purchase history information. RMA attempts to assess this by evaluating the amount of variability of the household specific parameters that can be explained by the demographic variables, as opposed to unobserved heterogeneity which is given by the square root of the diagonal elements of V_β. For each of the regression coefficients in β_h, the following measure is used in RMA:

$$\rho^2 = 1 - \frac{\text{Var(unobservable component)}}{\text{Var(Total)}}. \tag{3.26}$$

This measure ρ^2 is similar to the coefficient of determination utilized in classical regression: a large value implies that the demographic variables are useful in predicting that parameter. Using a WINBUGS program, under the same prior specifications given in RMA, we reproduce their Table 4 for the full information set with complete data, and the new result is shown in Table 3.3. We use 5000 iterations for burn-in and $20,000$ iterations (with convergence diagnostics checked) to obtain these results. Note that some values in Table 3.3 are different from those in RMA, who used an incorrect full conditional distribution in their computation. A noteworthy difference is that the ρ^2 of display and feature coefficients are not small, both greater than 0.5, and their corresponding unobservable heterogeneity is around 0.34, which is quite small. According to the criterion discussed in RMA (1996), the demographic variables do have some capability to explain the marketing mix sensitivities.

To explore the effect of setting aside some data for the validation set, we run the RMA benchmark model using the diffuse proper prior specifications as stated in RMA (same as *diffuse proper priors I*) on the calibration part of the full information set. The results are shown in Table 3.4. As seen, there is no real significant difference between Tables 3.3 and 3.4, which indicates that retaining part of the data for validation does not affect the model's performance in this case.

Table 3.5 presents the estimates of the brand-specific utility variances as measured by the diagonal components of Λ (the first element is fixed at 1 for identification purposes). The table shows that the variance estimates are very different among

Table 3.3. Empirical Analysis - Posterior estimates of Δ coefficients and the associated probabilities from the RMA benchmark model using diffuse proper priors I with complete data.

Beta	Cons	ln(Inc)	ln(Fam Size)	Retire	Unemp HH	Single Mom	Unobs. Hetero	ρ^2
Starkist Water Int	-0.08	0.12	-0.02	-0.04	**0.72**	-0.18	0.89	0.09
	[0.83]	(0.85)	[0.55]	[0.56]	**(0.99)**	[0.77]		
Private Label Int	**-4.61**	**-1.29**	0.33	-0.27	**2.61**	-0.45	2.69	0.28
	[1.00]	**[1.00]**	(0.77)	[0.63]	**(1.00)**	[0.74]		
C-O-S Oil Int	**-1.65**	**-0.25**	0.16	0.25	0.44	0.06	1.93	0.02
	[1.00]	**[0.90]**	(0.73)	(0.77)	(0.83)	(0.55)		
Starkist Oil Int	**-2.19**	-0.24	-0.09	0.42	**1.55**	-0.10	2.44	0.05
	[1.00]	[0.83]	[0.61]	(0.83)	**(0.99)**	[0.58]		
Price Coef	**-7.14**	-0.16	-0.49	**-1.81**	-0.85	-0.07	3.40	0.07
	[1.00]	[0.63]	[0.80]	**[0.98]**	[0.75]	[0.52]		
Display Coeff	**0.22**	-0.07	-0.07	**-0.36**	**-0.48**	0.08	0.32	0.67
	(1.00)	[0.72]	[0.68]	**[0.95]**	**[0.96]**	(0.65)		
Feature Coef	**0.37**	0.02	0.10	**0.40**	0.28	0.09	0.35	0.54
	(1.00)	(0.58)	(0.79)	**(0.99)**	(0.89)	(0.66)		

Note:
( ) indicates probability that the coefficient is positive.
[ ] indicates probability that the coefficient is negative.
Bold indicates probability exceeds 0.90.

Table 3.4. Empirical Analysis - Posterior estimates of Δ coefficients and the associated probabilities from the RMA benchmark model using diffuse proper priors I with the calibration set

Beta	Cons	ln(Inc)	ln(Fam Size)	Retire	Unemp HH	Single Mom	Unobs. Hetero	ρ^2
Starkist Water Int	0.004	0.07	-0.03	0.02	**0.64**	-0.11	0.76	0.10
	(0.53)	(0.73)	[0.58]	(0.53)	**(0.99)**	(0.70)		
Private Label Int	**-3.46**	**-1.09**	0.40	0.08	**1.98**	-0.19	2.00	0.33
	[1.00]	**[1.00]**	(0.85)	(0.56)	**(1.00)**	[0.63]		
C-O-S Oil Int	**-1.34**	-0.14	0.05	0.14	0.41	-0.01	1.69	0.01
	[1.00]	[0.77]	(0.58)	(0.67)	(0.83)	[0.52]		
Starkist Oil Int	**-1.80**	-0.19	-0.10	0.42	**1.32**	-0.35	2.10	0.05
	[1.00]	[0.78]	[0.63]	(0.85)	**(0.99)**	[0.76]		
Price Coef	**-6.24**	-0.38	-0.58	**-1.45**	-0.63	-0.24	2.90	0.08
	[1.00]	[0.83]	[0.86]	**[0.98]**	[0.75]	[0.61]		
Display Coeff	**0.18**	-0.02	-0.12	**-0.35**	**-0.49**	0.10	0.30	0.72
	(1.00)	[0.54]	[0.80]	**[0.95]**	**[0.98]**	(0.67)		
Feature Coef	**0.32**	0.05	0.04	**0.47**	0.16	0.09	0.30	0.71
	(1.00)	(0.71)	(0.62)	**(1.00)**	(0.78)	(0.68)		

Note:
( ) indicates probability that the coefficient is positive.
[ ] indicates probability that the coefficient is negative.
Bold indicates probability exceeds 0.90.

Table 3.5. Empirical Analysis - Posterior mean estimates of the diagonal components of Λ from the benchmark model using diffuse proper priors I.

Brand	Variance
C-O-S Water	1
Starkist Water	0.25
Private Label	0.52
C-O-S Oil	1.29
Starkist Oil	1.17

the brands. For example, the brand utility variance of Starkist water-packed (which equals 0.25) is just one-fifth that of C-O-S oil-packed (variance estimate is 1.29). Does the substantial variability among brand utility variances imply poor results from our proposed model? The answer is no based on our results to be shown below.

We apply our proposed model to the tuna calibration set using *diffuse proper priors I*, and the corresponding results are in Table 3.6. If we compare Table 3.6 with Table 3.4, it is seen that the results are, at first glance, somewhat similar even though we assume a correlation matrix in Equation (3.17). This occurs because, as noted before, only the relative utilities matter in the formulation, and the distribution of the relative utilities in our model is more general than that of RMA. Hence, reasonable results from the RMA model are likely to be obtained from our model as well.

Although the above tables show that our model and the RMA benchmark model can yield similar results, we argue that our model is better because it provides more stable estimates with respect to proper priors with large variances. Indeed, the posterior distribution from our model is always proper (its proof is given in Appendix C), but that is not the case with the RMA benchmark model. For example, the Bayesian estimates with respect to *diffuse proper priors II* from our proposed model and the RMA benchmark model are shown in Tables 3.7 and 3.8, respectively. Note that the results in Table 3.7 are similar to those in Table 3.6. However, in comparing Tables

Table 3.6. Empirical Analysis - Posterior estimates of Δ coefficients and the associated probabililties from our proposed model using diffuse proper priors I with the calibration set

Beta	Cons	ln(Inc)	ln(Fam Size)	Retire	Unemp HH	Single Mom	Unobs. Hetero	ρ^2
Starkist Water Int	-0.09	0.12	-0.04	0.02	**0.66**	-0.12	0.79	0.12
	[0.88]	(0.82)	[0.60]	(0.54)	**(0.98)**	[0.68]		
Private Label Int	**-4.14**	**-1.31**	0.47	0.20	**2.35**	-0.22	2.30	0.34
	[1.00]	**[1.00]**	(0.86)	(0.63)	**[1.00]**	[0.62]		
C-O-S Oil Int	**-1.33**	-0.16	0.05	0.13	0.41	-0.04	1.72	0.01
	[1.00]	[0.79]	(0.58)	(0.65)	(0.83)	[0.54]		
Starkist Oil Int	**-1.98**	-0.21	-0.15	0.43	**1.36**	-0.32	2.20	0.05
	[1.00]	[0.80]	[0.67]	(0.84)	**(0.99)**	[0.74]		
Price Coef	**-6.82**	-0.33	-0.69	**-1.33**	-0.71	-0.08	3.21	0.06
	[1.00]	[0.78]	[0.87]	**[0.94]**	[0.74]	[0.53]		
Display Coeff	**0.24**	-0.01	-0.09	**-0.42**	**-0.56**	0.13	0.31	0.80
	(1.00)	[0.53]	[0.70]	**[0.98]**	**[0.98]**	(0.69)		
Feature Coef	**0.33**	0.08	0.07	**0.53**	0.24	0.16	0.33	0.69
	(1.00)	(0.76)	(0.69)	**(1.00)**	(0.84)	(0.78)		

Note:
( ) indicates probability that the coefficient is positive.
[ ] indicates probability that the coefficient is negative.
Bold indicates probability exceeds 0.90.

3.8 and 3.4, we find that the results of the benchmark model with different prior distributions are substantially different. Indeed, results from the RMA benchmark model are sensitive to prior specifications, and, in an extreme case when only one observation is recorded per household, their posterior distribution can be shown to be improper in the limit. (Its proof is given in Appendix D.)

Furthermore, we compare the power of each model to predict household purchases in a holdout sample. We compare the predicted choices with the actual observations in the holdout data set, and compute the proportion of correct predictions. The corresponding results are shown in Table 3.9. When *diffuse proper priors I* are used, out-of-sample prediction accuracies from the two models are comparable. However, when *diffuse proper priors II* are used, our proposed model outperforms the RMA benchmark model substantially. We also observe that when the more vague prior is adopted, the prediction accuracy on future purchases of the RMA benchmark model

Table 3.7. Empirical Analysis - Posterior estimates of Δ coefficients and the associated probabilities from our proposed model using diffuse proper priors II with the calibration data set

Beta	Cons	ln(Inc)	ln(Fam Size)	Retire	Unemp HH	Single Mom	Unobs. Hetero	ρ^2
Starkist Water Int	-0.09	0.12	-0.04	0.02	**0.65**	-0.11	0.79	0.11
	[0.88]	(0.82)	[0.60]	(0.54)	**(0.98)**	[0.66]		
Private Label Int	**-4.18**	**-1.31**	0.45	0.17	**2.37**	-0.23	2.33	0.33
	[1.00]	**[1.00]**	(0.85)	(0.62)	**(1.00)**	[0.63]		
C-O-S Oil Int	**-1.33**	-0.15	0.06	0.12	0.41	-0.02	1.72	0.01
	[1.00]	[0.79]	(0.59)	(0.64)	(0.82)	[0.53]		
Starkist Oil Int	**-1.98**	-0.20	-0.16	0.43	**1.36**	-0.31	2.20	0.05
	[1.00]	[0.80]	[0.69]	(0.85)	**(0.99)**	[0.74]		
Price Coef	**-6.82**	-0.35	-0.70	**-1.27**	-0.65	-0.14	3.22	0.06
	[1.00]	[0.79]	[0.88]	**[0.95]**	[0.73]	[0.56]		
Display Coeff	**0.24**	-0.01	-0.09	**-0.44**	**-0.57**	0.12	0.31	0.81
	(1.00)	[0.52]	[0.71]	**[0.98]**	**[0.98]**	(0.68)		
Feature Coef	**0.33**	0.08	0.07	**0.54**	0.25	0.17	0.33	0.71
	(1.00)	(0.76)	(0.71)	**(1.00)**	(0.85)	(0.78)		

Note:
( ) indicates probability that the coefficient is positive.
[ ] indicates probability that the coefficient is negative.
Bold indicates probability exceeds 0.90.

Table 3.8. Empirical Analysis - Posterior estimates of Δ coefficients and the associated probabilities from the RMA benchmark model using diffuse proper priors II with the calibration data set.

Beta	Cons	ln(Inc)	ln(Fam Size)	Retire	Unemp HH	Single Mom	Unobs. Hetero	ρ^2
Starkist Water Int	**-1211.63**	**139.04**	-56.56	81.79	63.89	**-303.33**	906.52	0.03
	[1.00]	**(0.92)**	[0.66]	(0.69)	(0.61)	**[0.91]**		
Private Label Int	**-1253.04**	**-345.20**	**228.22**	-35.46	**530.93**	100.43	732.04	0.23
	[1.00]	**[1.00]**	**(0.96)**	[0.57]	**(1.00)**	(0.70)		
C-O-S Oil Int	**-20607.21**	264.00	453.47	2283.55	2682.34	-1741.82	10631.35	0.02
	[1.00]	(0.58)	(0.61)	(0.87)	(0.84)	[0.75]		
Starkist Oil Int	**-383.90**	36.90	-1.59	**84.73**	14.16	-27.50	368.24	0.02
	[1.00]	(0.82)	[0.51]	**(0.90)**	(0.56)	[0.63]		
Price Coef	**-614.77**	36.99	34.16	**179.23**	104.47	152.98	732.42	0.01
	[1.00]	(0.71)	(0.64)	**(0.94)**	(0.75)	(0.87)		
Display Coeff	**56.98**	40.52	27.39	88.08	85.34	92.33	572.09	0.01
	(0.96	(0.74)	(0.63)	(0.79)	(0.73)	(0.77)		
Feature Coef	17.47	8.73	**118.15**	66.92	**152.73**	**163.82**	394.26	0.04
	(0.79)	(0.59)	**(0.99)**	(0.86)	**(0.96)**	**(0.97)**		

Note:
( ) indicates probability that the coefficient is positive.
[ ] indicates probability that the coefficient is negative.
Bold indicates probability exceeds 0.90.

Table 3.9. Empirical Analysis - Holdout sample prediction accuracy comparison between the RMA benchmark model and our proposed model.

Priors Specification	Rossi et al. (1996)	Proposed Model
Diffuse Proper Prior I	76%	**77%**
Diffuse Proper Prior II	51%	**76%**

declines by 25%, which is a large drawback on the model performance.

3.6 Conclusion

Rossi et al. (1996) showed that there exists substantial potential for improving the profitability of direct marketing efforts by more fully utilizing household purchase histories. Managers can rely on customer purchasing data to formulate their micro-marketing strategies. Thus, accurate and stable parameter estimates are important factors in the decision making process. In this chapter, we identify an error in RMA, and show that our proposed model is less sensitive to prior specifications as compared to the RMA benchmark model, and thus it provides a more useful tool to the management. This is especially important as estimates from the RMA benchmark model vary dramatically depending on the priors chosen for the analysis. While an argument can be made to use informative priors in Bayesian inference with sparse data (Lenk & Orme, 2009), we note that diffuse proper priors are commonly used in practice. As situations to "let the data speak" are preferred and various diffuse proper priors can be assumed, a model that is less sensitive to prior specifications should give users more confidence in the final results of an analysis.

Our proposed model contributes to the literature in three aspects. First, it is a flexible and general model. By assuming a correlation matrix for the brand utilities, there is no loss of generality as it yields a general covariance structure on the relative utilities with unequal variances and unequal covariances without sign restrictions. The RMA benchmark model is restrictive because it assumes a positive equal covariance between relative utilities. Second, our model appears to be more robust than the RMA benchmark model with respect to prior specifications. We show that the posterior distribution of our model is always proper, while the RMA benchmark model becomes improper in the limit when there is only one observation per household and the variances of the prior distributions are set large to represent vague information. Lastly, although it is usually difficult to generate a correlation matrix due to the lack of a natural conjugate prior, we employ a parameter expansion technique to develop an efficient algorithm to sample correlation matrix. This algorithm is approximately Gibbs sampling and is more efficient than existing Metropolis-Hasting methods used to generate a covariance matrix with constraints.

Appendices

Appendix A. Derivation of the Full Conditional Distributions for Our Proposed Model

1. Proof of Equation (3.21):

   Since β_h denotes the reduced $(k-1)$-dimensional vector of parameters, Equation 3.17) becomes:

   $$y_{ht} = \tilde{X}_{ht}\beta_h + \varepsilon_{ht},$$

   where $\tilde{X}_{ht}$ is the reduced matrix of X_{ht} by deleting its first column. Then, the full conditional distribution of y_{ht} is:

   $\pi(y_{ht}|$ all others$)$

$$\propto \exp\{-\frac{1}{2}(y_{ht} - \tilde{X}_{ht}\beta_h)' R^{-1}(y_{ht} - \tilde{X}_{ht}\beta_h)\} I(y_{ht}[I_{ht}] > \max(y_{ht}[-I_{ht}])).$$

So $y_{ht}|$ all others follows a truncated normal distribution, TN($\tilde{X}_{ht}\beta_h, R$), where the truncation is such that $y_{ht}[I_{ht}] > \max(y_{ht}[-I_{ht}])$. Thus the j^{th} component of y_{ht}, $y_{ht}[j]$, has a univariate truncated normal distribution conditional on all other components of y_{ht} and other parameters. Let D_j be a matrix that switches the first and the j^{th} components of y_{ht}:

$$D_j \tilde{X}_{ht}\beta_h \triangleq \begin{pmatrix} \mu_j \\ \mu_{-j} \end{pmatrix}, \text{ and } D_j R D_j' \triangleq \begin{pmatrix} r_{jj} & \mathbf{r}_{j12} \\ \mathbf{r}_{j21} & \mathbf{r}_{j22} \end{pmatrix}.$$

Then,

$$y_{ht}[j]| \text{ all others} \sim TN(\mu_j + \mathbf{r}_{j12}\mathbf{r}_{j22}^{-1}(y_{ht}[-j] - \mu_{-j}), r_{jj} - \mathbf{r}_{j12}\mathbf{r}_{j22}^{-1}\mathbf{r}_{j21}),$$

where the truncation is given by:

$$\begin{cases} (\max(y_{ht}[-j]), +\infty) & \text{if } j = I_{ht}, \\ (-\infty, y_{ht}[I_{ht}]) & \text{if } j \neq I_{ht}. \end{cases}$$

2. Proof of Equation (3.22):

The full conditional distribution of β_h is:

$$\begin{aligned}
&\pi(\beta_h|\text{ all others})\\
&\propto \exp\Bigg\{-\frac{1}{2}\sum_{t=1}^{T_h}\Big[(y_{ht}-\tilde{X}_{ht}\beta_h)'R^{-1}(y_{ht}-\tilde{X}_{ht}\beta_h)\Big]\\
&\quad -\frac{1}{2}(\beta_h-\Delta Z_h)'V_\beta^{-1}(\beta_h-\Delta Z_h)\Bigg\}\\
&\propto \exp\Bigg\{-\frac{1}{2}\beta_h'\left(\sum_t(\tilde{X}_{ht}'R^{-1}\tilde{X}_{ht})+V_\beta^{-1}\right)\beta_h\\
&\quad +\beta_h'\left(\sum_t(\tilde{X}_{ht}'R^{-1}y_{ht})+V_\beta^{-1}\Delta Z_h\right)\Bigg\}.
\end{aligned}$$

So, $\beta_h|$ all others $\sim N(\beta_h^0, V_\beta^n)$, where:

$$V_\beta^n=\left(\sum_t(\tilde{X}_{ht}'R^{-1}\tilde{X}_{ht})+V_\beta^{-1}\right)^{-1} \text{ and } \beta_h^0=V_\beta^n\left(\sum_t(\tilde{X}_{ht}'R^{-1}y_{ht})+V_\beta^{-1}\Delta Z_h\right).$$

3. Proof of Equation (3.23):

   The full conditional distribution of Δ is:

$$\begin{aligned}
&\pi(\Delta|\text{ all others})\\
&\propto\text{etr}\Bigg\{-\frac{1}{2}V_\beta^{-1}\Bigg[(B-\Delta Z)(B-\Delta Z)'+(\Delta-\Delta_0)A_d(\Delta-\Delta_0)'\Bigg]\Bigg\}\\
&\propto\text{etr}\Bigg\{V_\beta^{-1}\Bigg[\Delta(ZZ'+A_d)\Delta'-2\Delta(ZB'+A_d\Delta_0')\Bigg]\Bigg\}.
\end{aligned}$$

   So, $\Delta|$ all others $\sim MN(\Delta_{0n}, V_\beta, (ZZ'+A_d)^{-1})$, where:

$$\Delta_{0n}=(BZ'+\Delta_0A_d)(ZZ'+A_d)^{-1}.$$

4. Proof of Equation (3.24):

The full conditional distribution of V_β^{-1} is:

$$
\begin{aligned}
&\pi(V_\beta^{-1}| \text{ all others}) \\
\propto & |V_\beta^{-1}|^{H/2} \text{etr}\left\{-\frac{1}{2}V_\beta^{-1}(B-\Delta Z)(B-\Delta Z)'\right\} \\
&\times |V_\beta^{-1}|^{l/2} \text{etr}\left\{-\frac{1}{2}V_\beta^{-1}(\Delta-\Delta_0)A_d(\Delta-\Delta_0)'\right\} \\
&\times |V_\beta^{-1}|^{\frac{v-k}{2}} \text{etr}\left\{-\frac{1}{2}V_\beta^{-1}V^{-1}\right\}.
\end{aligned}
$$

Note that RMA (1996) does not include the term, $|V_\beta^{-1}|^{l/2}\text{etr}\left\{-\frac{1}{2}V_\beta^{-1}(\Delta-\Delta_0)A_d(\Delta-\Delta_0)'\right\}$ in their derivation of the full conditional distribution, so the expression in their paper is incorrect. Then, our full conditional distribution is proportional to:

$$
|V_\beta^{-1}|^{\frac{v+H+l-k}{2}} \text{etr}\left\{V_\beta^{-1}\left[(B-\Delta Z)(B-\Delta Z)'+(\Delta-\Delta_0)A_d(\Delta-\Delta_0)'+V^{-1}\right]\right\}.
$$

So, $V_\beta^{-1}|$ all others $\sim W(v+H+l, V_n^{-1})$, where:

$$
V_n = (B-\Delta Z)(B-\Delta Z)'+(\Delta-\Delta_0)A_d(\Delta-\Delta_0)'+V^{-1}.
$$

Appendix B. Parameter Expansion Algorithm on Sampling Correlation Matrix

Stage I: Parameter Expanded Reparameterization

We want to transform R into a less constrained covariance matrix Σ through a parameter expansion scheme. For any Σ, we can find a unique R such that $\Sigma =$

$D^{1/2}RD^{1/2}$, where D is a diagonal matrix. Thus the mapping from Σ to R exists:

$$P: \ \ \Sigma \rightarrow D^{-1/2}\Sigma D^{-1/2} = R.$$

To find the mapping from R to Σ, we need to impose some constraints. Let $y^*_{ht} = D^{1/2}(y_{ht} - \tilde{X}_{ht}\beta_h)$, and we assume that $\sum_{h=1}^{H}\sum_{t=1}^{T_h}(y^*_{ht}[j])^2 = 1, \forall j = 1, ..., m$. Then, there exists a mapping from R to Σ:

$$P^{-1}: \ \ R \rightarrow D^{1/2}RD1/2 = \Sigma,$$

where $D^{1/2}_{jj} = \left[\sum_{h=1}^{H}\sum_{t=1}^{T_h}(y_{ht}[j] - \tilde{X}_{ht}[j]\beta_h)^2\right]^{-1/2}$ which is derived from the imposed constraint.

It is clear that this is a one-to-one mapping. Under the prior distribution of R as stated in Equation (3.20), the full conditional distribution of Σ is not a standard distribution. However, this can be achieved by employing the following candidate prior on R:

$$\pi(R) \propto |R|^{-(m+1)/2}.$$

The full conditional distribution of Σ is then:

$$\pi(\Sigma|\text{ all others}) \propto |\Sigma|^{-\frac{\sum_h T_h + m + 1}{2}} \exp\left\{-\frac{1}{2}\text{tr}(S\Sigma^{-1})\right\}.$$

So, $\Sigma|$ all others $\sim IW(\sum_h T_h, S)$, where $S = \sum_{h=1}^{H}\sum_{t=1}^{T_h}(y_{ht} - \tilde{X}_{ht}\beta_h)'D(y_{ht} - \tilde{X}_{ht}\beta_h)$.

Stage II: Parameter Expanded Metropolis-Hastings

After we have obtained a random deviate of Σ from the above Inverse Wishart distribution, we can transform it into R by using the mapping P, i.e., $R^* = D^{-1/2}\Sigma D^{-1/2}$. Since R^* is obtained based on the candidate prior, we use the Metropolis-Hasting algorithm to determine if this new random deviate should be accepted. We will accept R^* with probability:

$$\min\left\{1, \exp\left[\frac{m+1}{2}(log|R^*| - log|R^{(n)}|)\right]\right\},$$

where $R^{(n)}$ is the estimate of R at the previous iteration.

Appendix C: Proof of Propriety of the Posterior Distribution from the Proposed Model

We prove the result with one observation per household. Let $X = \left[\tilde{X}_1, \ldots, \tilde{X}_H\right]$, and $Y = [y_1', \ldots, y_N']'$. The likelihood function of our model is:

$$\begin{aligned} & L(Y, B, \Delta, R, V_\beta | I, X, Z) \\ \propto & \prod_{h=1}^{H} \left[I(y_h[I_h] > \max(y_h[-I_h]))\right] |R|^{-\frac{H}{2}} |V_\beta|^{-\frac{H}{2}} \\ & \times \exp\left\{-\frac{1}{2}\sum_{h=1}^{H}\Big((y_h - \tilde{X}_h\beta_h)' R^{-1}(y_h - \tilde{X}_h\beta_h) + (\beta_h - \Delta Z_h)' V_\beta^{-1}(\beta_h - \Delta Z_h)\Big)\right\}. \end{aligned}$$

The joint posterior distribution can be expressed as follows:

$$\begin{aligned} & \pi(Y, B, \Delta, R, V_\beta | I, X, Z) \\ = & \left[\prod_{h=1}^{H} \pi(\beta_h | I, X, Z, y_h, \Delta, R, V_\beta)\right] \pi(\Delta | I, X, Z, Y, R, V_\beta) \\ & \times \pi(Y | I, X, Z, R, V_\beta) \pi(R | I, X, Z, V_\beta) \pi(V_\beta | I, X, Z). \end{aligned}$$

The joint posterior distribution is a proper probability distribution if each (conditional) posterior distribution on the right hand side of the above equation is proper.

- For $h = 1, \ldots, H$,

$$\pi(\beta_h|I, X, Z, Y, \Delta, R, V_\beta) = \pi(\beta_h|X, Z, Y, \Delta, R, V_\beta)$$
$$\propto \exp\left\{-\frac{1}{2}\left[(y_h - \tilde{X}_h\beta_h)'R^{-1}(y_h - \tilde{X}_h\beta_h) + (\beta_h - \Delta Z_h)'V_\beta^{-1}(\beta_h - \Delta Z_h)\right]\right\}.$$

So, $\beta_h\Big|I, X, Z, Y, \Delta, R, V_\beta \sim N(b_h^0, V_\beta^{\text{new}})$, where:

$$V_\beta^{\text{new}} = \left(\tilde{X}_h'R^{-1}\tilde{X}_h + V_\beta^{-1}\right)^{-1} \text{ and } b_h^0 = V_\beta^{\text{new}}\left(\tilde{X}_h'R^{-1}Y_h + V_\beta^{-1}\Delta Z_h\right).$$

Thus, the conditional posterior distribution of $\beta_h|I, X, Z, Y, \Delta, R, V_\beta$ is proper.

- Let η be the vectorization of Δ, and $\Upsilon_h = \tilde{X}_h(E_{k-1} \bigotimes Z_h)'$, where E_{k-1} is the identity matrix with dimension $(k-1) \times (k-1)$ and $\bigotimes$ denotes the Kronecker product. $\pi(\Delta|I, X, Z, Y, R, V_\beta)$ is equivalent to $\pi(\eta|I, X, Z, Y, R, V_\beta)$ where:

$$\pi(\eta|I, X, Z, Y, R, V_\beta) = \pi(\eta|X, Z, Y, R, V_\beta) \propto L(\eta|X, Z, Y, R, V_\beta)\pi(\eta|V_\beta)$$
$$\propto \exp\left\{-\frac{1}{2}\sum_{h=1}^{H}\left[(y_h - \Upsilon_h\eta)'(\tilde{X}_hV_\beta\tilde{X}_h' + R)^{-1}(y_h - \Upsilon_h\eta)\right]\right\}$$
$$\times \exp\left\{-\frac{1}{2}(\eta - \bar{d})'(V_\beta \bigotimes A_d^{-1})^{-1}(\eta - \bar{d})\right\}$$
$$\propto \exp\Bigg\{-\frac{1}{2}\eta'\left((V_\beta^{-1} \bigotimes A_d) + \sum_{h=1}^{H}[\Upsilon_h'(\tilde{X}_hV_\beta\tilde{X}_h' + R)^{-1}\Upsilon_h]\right)\eta$$
$$+ \eta'\left(\sum_{h=1}^{H}[\Upsilon_h'(\tilde{X}_hV_\beta\tilde{X}_h' + R)^{-1}y_h] + (V_\beta^{-1} \bigotimes A_d)\bar{d}\right)\Bigg\}.$$

Typically, one lets A_d in the prior distribution to approach a zero matrix to represent vague information. Then, in the limit, the above expression goes to:

$$\exp\Bigg\{ -\frac{1}{2}\eta' \left(\sum_{h=1}^{H} [\Upsilon_h' (\tilde{X}_h V_\beta \tilde{X}_h' + R)^{-1} \Upsilon_h] \right) \eta + \eta' \left(\sum_{h=1}^{H} [\Upsilon_h' (\tilde{X}_h V_\beta \tilde{X}_h' + R)^{-1} y_h] \right) \Bigg\}.$$

Thus, $\eta | I, X, Z, Y, R, V_\beta \sim N(V_\eta d_0, V_\eta)$, where:

$$V_\eta = \left[\sum_{h=1}^{H} \left(\Upsilon_h' (\tilde{X}_h V_\beta \tilde{X}_h' + R)^{-1} \Upsilon_h \right) \right]^{-1},$$
$$d_0 = \sum_{h=1}^{H} \left(\Upsilon_h' (\tilde{X}_h V_\beta \tilde{X}_h' + R)^{-1} y_h \right).$$

So, the conditional posterior distribution of $\eta | I, X, Z, Y, R, V_\beta$ is proper.

- Let

$$\Upsilon = \begin{pmatrix} \Upsilon_1 \\ \vdots \\ \Upsilon_H \end{pmatrix} \text{ and } \Psi = \begin{pmatrix} (\tilde{X}_1 V_\beta \tilde{X}_1' + R)^{-1} & & \\ & \ddots & \\ & & (\tilde{X}_H V_\beta \tilde{X}_H' + R)^{-1} \end{pmatrix}.$$

Since

$$\pi(Y | I, X, Z, R, V_\beta) \propto \exp\left\{ -\frac{1}{2} Y' \left(\Psi - \Psi\Upsilon(\Upsilon'\Psi\Upsilon)^{-1}\Upsilon'\Psi \right) Y \right\} \times \prod_h \prod_t \left[I\{ y_{ht}[I_{ht}] > \max(y_{ht}[-I_{ht}]) \} \right].$$

So, $Y | I, X, Z, R, V_\beta \sim TN(0, (\Psi - \Psi\Upsilon(\Upsilon'\Psi\Upsilon)^{-1}\Upsilon'\Psi)^{-1})$ with truncation restriction on $I\{y_{ht}[I_{ht}] > y_{ht}[-I_{ht}]\}$ for any $h = 1, \ldots, N$ and $t = 1, \ldots, T_h$.

Thus, the conditional posterior distribution of $Y|I, X, Z, R, V_\beta$ is proper.

- The space of m-dimension correlation matrix is convex and compact (Rousseeuw & Molenberghs, 1994), so the prior distribution we specify on R is a proper distribution. Thus, the conditional posterior distribution $\pi(R|I, X, Z, V_\beta)$ is also proper.

- Similar to the argument given above, because the prior distribution on V_β^{-1} is Wishart which is a proper distribution, the conditional distribution of $V_\beta|I, X, Z$ is proper.

Appendix D: Proof of Impropriety of the Posterior Distribution from the RMA Benchmark Model with One Observation per Household

Recall that $\Lambda = diag(1, \sigma_2^2, \ldots, \sigma_m^2)$, we let:

$$\Psi^* = \begin{pmatrix} (\tilde{X}_1 V_\beta \tilde{X}_1' + \Lambda)^{-1} & & \\ & \ddots & \\ & & (\tilde{X}_H V_\beta \tilde{X}_H' + \Lambda)^{-1} \end{pmatrix}.$$

Then, we have:

$$\begin{aligned} & \pi(\Lambda|I, X, Z, Y, V_\beta) \\ =& C_1 \prod_{j=2}^{m} \left[(\sigma_j^{-2})^{\nu-1} \exp\{-\frac{\sigma_j^{-2}}{v_j}\} \right] \\ & \times \left(\frac{|\Psi^*|}{|\Upsilon' \Psi^* \Upsilon|} \right)^{\frac{1}{2}} \exp\left\{ -\frac{1}{2} Y' \left(\Psi^* - \Psi^* \Upsilon (\Upsilon' \Psi^* \Upsilon)^{-1} \Upsilon' \Psi^* \right) Y \right\}, \end{aligned}$$

where C_1 is the normalizing constant. Again, we set the variance of the Gamma distribution to be large to represent vague information. So, in the limit as we let $\nu \to 0$ and $\upsilon_j \to \infty$, the above conditional posterior distribution becomes:

$$\begin{aligned}
&\pi(\Lambda|I, X, Z, Y, V_\beta) \\
=&C_1\Big[\prod_{j=2}^{m}(\sigma_j^{-2})^{-1}\Big]\left(\frac{|\Psi^*|}{|\Upsilon'\Psi^*\Upsilon|}\right)^{\frac{1}{2}} \exp\Big\{-\frac{1}{2}Y'\Big(\Psi^* - \Psi^*\Upsilon(\Upsilon'\Psi^*\Upsilon)^{-1}\Upsilon'\Psi^*\Big)Y\Big\}.
\end{aligned}$$

Let:

$$\begin{aligned}
f(\sigma^{-2}) &= \prod_{j=2}^{m}(\sigma_j^{-2})^{-1} \\
g(\sigma^{-2}) &= \left(\frac{|\Psi^*|}{|\Upsilon'\Psi^*\Upsilon|}\right)^{\frac{1}{2}} exp\Big\{-\frac{1}{2}Y'\Big(\Psi^* - \Psi^*\Upsilon(\Upsilon'\Psi^*\Upsilon)^{-1}\Upsilon'\Psi^*\Big)Y\Big\}.
\end{aligned}$$

Then,

$$\begin{aligned}
&\int \pi(\Lambda|I, X, Z, Y, V_\beta)d\Lambda \\
=&\int_0^\infty \cdots \int_0^\infty \pi(\sigma_2^{-2}, \cdots, \sigma_m^{-2}|I, X, Z, Y, V_\beta)d\sigma_2^{-2}\cdots d\sigma_m^{-2} \\
\geq&\int_w^\infty \cdots \int_w^\infty \pi(\sigma_2^{-2}, \cdots, \sigma_m^{-2}|I, X, Z, Y, V_\beta)d\sigma_2^{-2}\cdots d\sigma_m^{-2}, \qquad (\forall w > 0) \\
=&C_1\int_w^\infty \cdots \int_w^\infty f(\sigma^{-2})g(\sigma^{-2})d\sigma_2^{-2}\cdots d\sigma_m^{-2}.
\end{aligned}$$

Let $\mathbf{F}$ be the region defined by $\{\sigma_j^{-2} > w, \forall j = 2, \ldots, m\}$. It is obvious that $g(\sigma^{-2}) > 0$ over $\mathbf{F}$ with the only possible exception when σ_j^{-2} approaches infinity. However, when $\sigma_j^{-2} \to \infty (j = 2, \ldots, m)$, $(\tilde{X}_j V_\beta \tilde{X}_j' + \Lambda)^{-1} \to (\tilde{X}_j V_\beta \tilde{X}_j' + \ddot{E})^{-1}$, where $\ddot{E}$ is a

$m \times m$ zero matrix except that its $(1,1)$ component is 1. Then,

$$\Psi^* \to \begin{pmatrix} (\tilde{X}_1 V_\beta \tilde{X}_1' + \ddot{E})^{-1} & & \\ & \ddots & \\ & & (\tilde{X}_H V_\beta \tilde{X}_H' + \ddot{E})^{-1} \end{pmatrix} \triangleq \Psi^*_\infty.$$

Therefore,

$$g(\sigma^{-2}) \to \left(\frac{|\Psi^*_\infty|}{|\Upsilon' \Psi^*_\infty \Upsilon|} \right)^{\frac{1}{2}} exp\left\{ -\frac{1}{2} Y' \left(\Psi^*_\infty - \Psi^*_\infty \Upsilon (\Upsilon' \Psi^*_\infty \Upsilon)^{-1} \Upsilon' \Psi^*_\infty \right) Y \right\} > 0.$$

Hence, $min_{\sigma^{-2} \in \mathbf{F}} g(\sigma^{-2}) = m_0$, where $0 < m_0 < \infty$. Thus:

$$\begin{aligned} \int \pi(\Lambda | I, X, Z, Y, V_\beta) d\Lambda \geq & C_1 \int_w^\infty \cdots \int_w^\infty f(\sigma^{-2}) g(\sigma^{-2}) d\sigma_2^{-2} \cdots d\sigma_m^{-2} \\ \geq & m_0 C_1 \int_w^\infty \cdots \int_w^\infty \left[\prod_{j=2}^m (\sigma_j^{-2})^{-1} \right] d\sigma_2^{-2} \cdots d\sigma_m^{-2} \\ = & m_0 C_1 \prod_{j=2}^m \left[\ln \sigma_j^{-2} \Big|_w^\infty \right] = \infty. \end{aligned}$$

Because $\int \pi(\Lambda | I, X, Z, Y, V_\beta) d\Lambda$ does not exist, $\pi(\Lambda | I, X, Z, Y, V_\beta)$ is <u>not</u> a proper distribution, and so the joint posterior distribution is improper.

Chapter 4

A Bayesian Vector Multidimensional Scaling with Variable Selection for the Analysis of Metric Rating Data

4.1 Introduction

Multidimensional scaling (MDS) is often used in marketing to produce a joint space map of consumers and brands in a reduced dimensionality. Various questions of managerial interest may be answered through such analysis. For example, what are the competing brand(s) for one specific brand? Who are the potential customers for a particular brand? How many underlying dimensions will influence consumers when they are forming purchasing decision, and what are those dimensions? Indeed, MDS is an important tool in product positioning as well for making strategic decisions.

Classical MDS procedures for analyzing two-way data and providing joint space representations of row and column entities/objects are classified into three major categories: unfolding representations (Takane, Young, & Leeuw, 1977), vector or scalar products representations (Carroll & Chang, 1970; Gifi, 1990) and optimal scaling or correspondence analysis (Benzécri, 1992; Shin & Fong, 1998). Although the classical methods are relatively easy to implement, they have a number of limitations. First, there lacks a solid basis for determining the dimensionality underlying the observed data. Classical deterministic MDS approaches suggest using ad-hoc scree plots examining stress or variance accounted for versus the dimensionality to find the optimal dimensionality. Second, identification constraints are frequently imposed in these models which can affect the derived solutions. Typically point estimates are obtained and it is difficult to assess the variability associated with the estimates. Third, it is hard, if not impossible, to decide which brand attributes affect the brand positions and the magnitude of impact when an effect exists.

To overcome these limitations, some scholars began constructing MDS models in a Bayesian framework. Oh and Raftery (2001) devise a Bayesian ideal point MDS method to analyze data which involves similarity or dissimilarity measures between pairs of objects. Martin and Quinn (2002) employ a dynamic Bayesian model to provide ideal point for all justices serving on the U.S. Supreme Court from 1953 to 1999. In their paper, they degenerate the dimension to a unidimensional case. Gormley and Murphy (2006) use a latent space model to analyze Irish election ranking data. They use a Procrustes method to solve the identification problem including rotation and translations. DeSarbo et al. (1999) present a Bayesian vector MDS (BVMDS) procedure for the spatial analysis of binary choice data. Fong et al. (2010) suggest a BVMDS model for the analysis of ordered successive categories preference/dominance data, and employ an informative prior distribution on the brand coordinates based on the posterior distribution derived from a related data set. DeSarbo et al. (2011) use a stochastic ordered preference BVMDS model to analyze ordered successive category measurements. In addition, Park et al. (2008) suggest a Bayesian MDS method that combines both the vector model and the ideal point model in a generalized framework for modeling metric dominance data. In a Bayesian framework, one can use the well-established model comparison criteria, including Bayes factor and DIC, to compare model fittings for each dimension, and avoid making ad-hoc decisions in determining the optimal dimensionality. Moreover, Bayesian analysis typically provides interval estimates which better account for the parameters' variability.

In addition, the Bayesian approach allows the incorporation of related data in interpreting each dimension. For existing classical models, the interpretation of the dimensions typically relies on a post-hoc property fitting analysis and the feature variables used to explain each dimension are not utilized in the model construction. However, it is often more reasonable to include such information directly in the statistical analysis, not after that for interpretation only. Fong et al. (2010) consider a related data set when creating an informative prior on the brand coordinates; however, they do not provide an explicit estimation on the impact of the brand features on the brand positions. In this chapter, we will develop a new BVMDS model that links the brand features with the brand coordinates directly.

In practice, there are many brand features associated with a specific brand, while the number of brands in the market is relatively smaller. As is well known, it is difficult to perform statistical inference with so many explanatory variables and only a small number of observations, the so called "large p, small n" problems described in the literature (Bernardo & Smith, 1994; George & McCulloch, 1997; O'Hara & Sillanp, 2009; Yi, George, & Allison, 2003; Meuwissen & Goddard, 2004). In our model, we incorporate a variable selection procedure focusing on the probabilistic fit in the form of latent mixture modeling to determine the optimal set of significant brand features contributing to brand positions and explaining the dimensionality.

This chapter is organized as follows. Section 4.2 provides a review of the classical and Bayesian MDS models, and discusses the pros and cons of the approaches. We then describe the proposed Bayesian vector multidimensional scaling with variable selection procedure in Section 4.3. We discuss how the model is set up, how to solve the identification problems associated with the vector MDS model, and propose a sampling algorithm to generate random deviates from the joint posterior distributions efficiently. Section 4.4 presents the results of two simulation studies comparing the proposed model with two competing benchmark models. Section 4.5 gives an empirical application analyzing large SUV purchase data. Finally, we summarize our contributions and provide a conclusion in Section 4.6.

4.2 Literature Review

4.2.1 Classical MDS Procedures

There are three classes of classical two-way MDS procedures to analyze two-way rating data: unfolding representations, vector or scalar product representations and correspondence analysis (Borg & Groenen, 2005). As we are proposing a new Bayesian vector model in this chapter, we will simply review the classical vector MDS model.

4.2.1.1 Vector or Scalar Product Model

Let Z_{ij} be the rating of consumer i on brand j, where a larger Z_{ij} indicates that the consumer prefers the brand more. Denoting a_i and b_j as the consumer and the brand vectors from the origin of a T-dimensional space, respectively, the basic idea of a vector model is to minimize the following stress function over all a_i and b_j:

$$\sigma_{\text{vector}}(A, B) = \sum_i \sum_j \left(Z_{ij} - a_i' b_j\right)^2,$$

A popular method to solve this problem is MDPREF (Carroll & Chang, 1970), which employs SVD on Z.

4.2.1.2 Discussion on the Classical MDS Approaches

An important issue with the classical vector MDS model concerns the choice of optimal dimensionality underlying the observed data. As one of the main goals of MDS, researchers want to use a reduced dimensionality to represent the data which needs to be balanced with better goodness-of-fit results from employing more dimensions. A popular procedure to determine the optimal number of dimensions is to do a scree test (Cattell, 1966) by plotting the fit value versus their corresponding dimensions, and choose the dimension where the smooth increase of the goodness-of-fit value appears to level off to the right of the plot. However, we argue that there is little basis for determining the optimal dimensionality for this methods. Sometimes, it is difficult to find the elbow of a stress plot. Other alternative method is to use information heuristics for ML-based MDS methods.

Attention should also be paid to the identification issue related to vector MDS models. The vector model suffers from identification problems including translation, rotation, reflection, and expansion/shrinkage relative to the origin. More specifically, for any solution (A, B) and a non-singular matrix M, if we set $\tilde{A} = MA$ and $\tilde{B} = MB$, where M is an orthogonal transformation such that $A'B = \tilde{A}'\tilde{B}$, then $(\tilde{A}, \tilde{B})$ will also be a solution. Typically, such identification problems are solved by a post-hoc Procrustes rotations applied on the parameter estimates from the classical methods.

4.2.2 Bayesian MDS Procedures

There has been a plethora of literature on developing MDS methodologies in a Bayesian framework, including the unfolding models (also called ideal-point models) (Oh & Raftery, 2001; Martin & Quinn, 2002; Gormley & Murphy, 2006), the vector models (DeSarbo, Kim, & Fong, 1999; Fong, DeSarbo, Park, & Scott, 2010; DeSarbo, Park, & Rao, 2011), and a mixture of these two models (Park, DeSarbo, & Liechty, 2008). As we discussed in Section 4.2.1.2, the optimal dimensionality is sometimes fuzzy based on the scree test. However, there is some well-established criteria in Bayesian theory that we can use to determine the optimal dimensionality. Moreover, the classical approaches typically solve the identification problems associated with an MDS model by doing a Procrustes transformation. The Bayesian approach offers more alternatives in that, as an alternative to the post-hoc transformation, imposing identification constraints or informative priors can also be done to solve the identification problems.

4.2.2.1 Existing Work on Bayesian MDS Models

Oh and Raftery (2001) propose a Bayesian metric MDS model to handle data involving similarity or dissimilarity measures between a pair of objects. In their work, they employ a new Bayesian criterion, which is based on the Bayes factor, to determine the dimension. The model is utilized to analyze a data set on the airline distances between cities, and the appropriateness of the new criterion they propose for choosing dimensionality is validated. Martin and Quinn (2002) employ a dynamic ideal point model for all justices serving on the U.S. Supreme Court from 1953 to 1999. They degenerate the dimension and only consider a unidimensional case, and find out that many justices have temporally changing ideal points. Gormley and Murphy (2006) use a latent space model to analyze Irish election ranking data so as to provide a joint space presentation of the candidates and voters in the same $T-$dimensional space. They use a Procrustes method to eradicate the identification problem including rotation and translations.

DeSarbo et al. (1999) present a Bayesian vector MDS procedure for the spatial analysis of binary choice data. They evade the identification problems by imposing strong prior on the brand coordinates. Fong et al. (2010) suggest a Bayesian vector MDS model for the analysis of ordered successive categories preference/dominance data. Instead of using conjugate prior on brand vectors, they make use of an external information in the form of an intractable posterior distribution derived from a related data set as the prior to estimate the spatial representation. More specifically, they assume a Bayesian factor model for the related data set, and the posterior of the factors is used as the prior distribution for the brand vector. DeSarbo et al. (2011) raise a stochastic ordered preference MDS vector model to analyze the ordered successive category measurement. Proper priors are specified in the model to circumvent indeterminacy problems.

Park et al. (2008) suggest a Bayesian MDS method that combines both the vector model and the ideal point model in a generalized framework for modeling metric dominance data. By doing this, they can account for both preference heterogeneity and structural heterogeneity. They impose constraints to their model to remove various types of identification problems associated with both the ideal point model and the vector model.

4.2.2.2 Discussion on the Existing Bayesian MDS Approaches

Obviously, there are many good features of fitting the MDS model in a Bayesian framework. We have more strategies to deal with different identification problems.

Also, there are several well-established methods that can be used to compare models. However, more efforts can be made to improve the existing models to address various research questions. For example, Fong et al. (2010) utilize a related data set to derive the informative prior distribution on the brand coordinates; a natural follow-up question is, "Can we use such information directly in the model construction?" It will be of great value if we can obtain information on which brand characteristics/features affect the brand positions, and how is the strength of such influence. Furthermore, there may be a lot of brand characteristics and features which are not helpful in assessing the brand coordinates. Then the second question is, "What are the significant features of a brand that determine the brand location?" To answer these interesting questions, we propose a new Bayesian vector MDS with variable selection procedure in the next section.

4.3 The Proposed Model

In this section, we develop a vector MDS model in a Bayesian framework to analyze consumer rating data. The model makes appropriate assumptions to solve the identification problems typically associated with a vector MDS model, and incorporate a variable selection procedure to identify the significant attributes. A sampling algorithm is proposed to generate random deviates from the joint posterior distribution efficiently.

4.3.1 Basic Model Form

We observe the metric preference rating response Z_{ij} from consumer i for brand j, where $i = 1, \ldots, N$ and $j = 1, \ldots, J$. We assume a vector model on Z_{ij} as:

$$Z_{ij} = a_i^{'} b_j + \varepsilon_{ij}, \tag{4.1}$$

where the error terms are independent and $\varepsilon_{ij} \sim N(0, \sigma^2)$. Here the $T-$dimensional vector a_i indicates the consumer vector, and the $T-$dimensional vector b_j indicates the brand vector. These vectors are assumed to be random and will be assigned appropriate prior distributions as in a Bayesian framework. Note that the model in Equation (4.1) is under-identified. Assuming that M is an orthogonal $T \times T$ matrix, then $\tilde{a}_i = Ma_i$ and $\tilde{b}_j = Mb_j$ is another solution to Equation (4.1) because $\tilde{a}_i^{'}\tilde{b}_j = a_i^{'} b_j$, where M is orthogonal. We will discuss how to solve these identification issues in Section 4.3.2.

Typically, we assume that consumer vectors independently follow a multivariate normal distribution as:

$$a_i \sim N(a_0, \Lambda), \quad \text{independently}, \tag{4.2}$$

where a_0 is the common mean, and Λ is a $T \times T$ positive definite matrix. On the other hand, if we have access to a related data set that contains brand characteristics and attributes, it will be reasonable to make use of such information to estimate the brand vectors. In our model, we adopt a multivariate normal regression model on b_j as:

$$b_j \sim N(b_0 + \Theta X_j, \Sigma), \quad \text{independently}, \tag{4.3}$$

where $X = (X_1, \ldots, X_J)$ is a $K \times J$ matrix that contains K observed brand covariates,

$T-$dimensional vector b_0 is a common brand intercept, $T \times K$ matrix Θ evaluates the relationship between the brand attributes and the brand vectors, and Σ is a $T \times T$ positive definite covariance matrix.

We employ conjugate priors as shown below:

$$\sigma^{-2} \sim \text{Ga}(\frac{m}{2}, \frac{m}{2}), \tag{4.4}$$

$$a_0 \sim N(0, G_a), \tag{4.5}$$

$$b_0 \sim N(0, G_b), \tag{4.6}$$

$$\Theta|\Sigma \sim \text{MN}(\Theta_0, \Sigma, H), \tag{4.7}$$

$$\Sigma^{-1} \sim W(\nu, V), \tag{4.8}$$

where the expectation of σ^{-2} is 1, $\text{MN}(\Theta_0, \Sigma, H)$ denotes a matrix normal distribution with mean Θ_0, row covariance matrix Σ, and column covariance matrix H (Dawid, 1981), and the expectation of Σ^{-1} is νV. G_a, G_b, Θ_0, ν and V are hyper-parameters that will be pre-specified. We will discuss the assumption on Λ in Section 4.3.2, and H in Section 4.3.3.

4.3.2 Identification Issues

A vector MDS model typically suffers from four identification problems: rotation, reflection, permutation and expansion/shrinkage relative to the origin. The first three problems are associated with the multiplication with an orthogonal matrix, and the last one can be achieved by multiplying the solution by a diagonal matrix with diagonal elements different from ± 1. As mentioned, an MDS model can solve these problems by three means: imposing identification constraints on the model, utilizing strong priors, or doing a Procrustes transformation. A Procrustes transformation is a geometric transformation that involves only reflection, rotation, uniform scaling, or a combination of these transformations. Therefore, it will not change the shape of the plot. In our model, we involve some constraints and incorporate a prior with certain structure to circumvent the indeterminacy problem. Details are given as follows:

- **Expansion/Shrinkage Problem:**

  We assume the covariance matrix of a_i to be an identity matrix multiplied by a pre-specified constant, i.e., $\Lambda = cI_T$, where c is a scalar, and I_T represents a T-dimensional identity matrix. Assuming that D is a diagonal matrix with some diagonal elements not equal to ± 1, and $Da_i \sim N(Da_0, D\Lambda D' = cDD)$. Because at least one of the diagonal components of D does not equal ± 1, $DD \neq I_T$. Therefore, D can only be a diagonal matrix with all diagonal components being ± 1, and we successfully rule out the expansion/shrinkage problem by assuming $\Lambda = cI_T$.

- **Rotation, Reflection and Permutation Problems:**

  Denote $A = (a_1, \ldots, a_N)$, and we assume that A is an upper triangular matrix with positive diagonal elements. Let Γ be an orthogonal matrix; if ΓA is upper triangular, Γ can only be a diagonal matrix with ± 1 on its diagonal. Furthermore, since all the diagonal elements of both A and ΓA should be positive, Γ can only be an identity matrix. Hence, there will be no rotation, reflection or permutation problem if A is an upper triangular matrix with positive diagonal elements. We call this constraint *the identifiable constraint.*

According to our assumptions above, Equation (4.2) can be rewritten as:

$$A \sim \mathrm{MN}(A_0, cI_T, I_N) I(A \text{ satisfies the identifiable constraint}), \tag{4.9}$$

where A_0 is a $T \times N$ matrix with each column equal to a_0. This is a constrained matrix normal distribution, which cannot be easily generated. An alternative sampling algorithm is proposed in Section 4.3.4 to relieve the intensive computing load caused by such constraint.

4.3.3 Model Choice and Variable Selection Procedure

The use of brand attributes to help estimate the brand parameters, as shown in Equation (4.3), is reasonable and beneficial. Instead of assuming a common intercept alone, the incorporation of data on brand-specific characteristics and features accommodates brand heterogeneity more directly. When linking brand position with the corresponding attributes, a popular and easy-to-interpret model is regression; however, an indeterminacy problem in the estimation of regression coefficients will arise when the number of attributes is greater than the number of available brands, which, in our case, is when $K + 1 > J$.

To solve this problem, our model incorporates a model and variable selection procedure (Brown, Vannucci, & Fearn, 1998). This variable selection method is based on probabilistic-fit in the form of latent mixture modeling. The basic idea is to assume a latent binary variable indicating whether the magnitude of the corresponding attribute is "large" or is concentrated on zero. In an extreme case when this variance is zero, the attributes with their corresponding indicator variable being equal to zero will be effectively deleted from the model. The value of the binary vector is evaluated in each iteration based on probabilistic-fit. For this extreme case, the deleted attributes will not be used for the estimation on the other variables in the next iteration. Therefore, we only use the "large" variables in each iteration of our MCMC process.

If we let $B = (b_1, \ldots, b_J)$, Equation (4.3) can be reformulated as:

$$B \sim \text{MN}(B_0 + \Theta X, \Sigma, I_J), \tag{4.10}$$

where B_0 is a $T \times J$ matrix with each column equal to b_0.

Now, we define a latent indicator vector $\gamma = (\gamma_k)$ with length K. If $\gamma_k = 1$, it indicates that the k^{th} attribute significantly contributes to the brand vectors B, while if $\gamma_k = 0$, it does not. If we only want to use the significant variables in each iteration, a strategy is to set the regression coefficient of the non-significant attribute very small, or even zero. So, we modify the prior on Θ to be:

$$\Theta \sim \text{MN}(\Theta_0, \Sigma, H_\gamma), \tag{4.11}$$

where H_γ is a diagonal matrix, and the k^{th} diagonal element of H_γ, h_{kk}, is defined as follows:

$$h_{kk} = \begin{cases} v_1 & \text{if } \gamma_k = 1; \\ v_0 & \text{if } \gamma_k = 0, \end{cases} \tag{4.12}$$

where $v_1 \gg v_0 \geq 0$. When Θ_0 is the zero matrix, it is of more interest to consider a "selection" prior, that is, we let $v_0 = 0$. (We will use this setting in the remaining part of this chapter.) Then, $\gamma_k = 0$ indicates that the k^{th} column of Θ has variance zero and mean zero. Since it is more appropriate to have a positive definite covariance matrix for a normal distribution, we rewrite the modified prior on Θ as:

$$\Theta_{(\gamma)} \sim \text{MN}(0, \Sigma, H_{(\gamma)}), \tag{4.13}$$

where $\Theta_{(\gamma)}$ is the degenerated matrix that selects the columns of Θ that have $\gamma_k = 1$, and $H_{(\gamma)} = v_1 I_l$, where $l = \sum_k \gamma_k$. If we also let $X_{(\gamma)}$ to contain the rows of X that have $\gamma_k = 1$, we will find that $\Theta X = \Theta_{(\gamma)} X_{(\gamma)}$ since the columns of Θ that have $\gamma_k = 0$ are all zero. Therefore, Equation (4.10) does not change after the degeneration of Θ.

Because γ is a binary indicator vector, a conjugate prior on γ_k is a Bernoulli distribution:

$$\gamma_k | w \sim \text{Ber}(w), \tag{4.14}$$

$$w \sim \text{Beta}(p, q), \tag{4.15}$$

where p and q are pre-specified. Then, under the model as shown in Equation (4.10) and the priors in Equation (4.6), (4.8), (4.13) and (4.14), when a large G_b is utilized, we have the conditional posterior distribution of γ as:

$$\pi(\gamma | B, w) \propto (|H_{(\gamma)}||K_{(\gamma)}|)^{-T/2} |V_{(\gamma)}|^{-(J+\nu+T-1)/2} \prod_k \gamma_k^w (1-\gamma_k)^{1-w}, \tag{4.16}$$

where:

$$K_{(\gamma)} = X_{(\gamma)} X_{(\gamma)}^{'} + H_{(\gamma)}^{-1},$$
$$V_{(\gamma)} = V^{-1} + BB^{'} - BX_{(\gamma)}^{'} K_{(\gamma)}^{-1} X_{(\gamma)} B^{'}.$$

For each iteration within MCMC, we obtain an estimated indicator vector γ with length K. Note that each component of this vector is either 1 or 0, so in total there are 2^K possible γ. At the end of the MCMC procedure, we count the frequency of each possibility, and the one with the highest frequency, that is the highest posterior probability, is declared the indicator vector of *the optimal model*.

4.3.4 The Proposed MCMC Sampling Procedure

As discussed in Section 4.3.2, we assume that the consumer vector matrix A follows the identifiable constraint to avoid the indeterminacy problems. A consequent prior on A is a matrix normal prior with constraint as shown in Equation (4.9). If we let $a_i^{P1} = (a_{i1}, \ldots, a_{ii})'$ and $a_i^{P2} = (a_{i,i+1}, \ldots, a_{iT})'$ for $i = 1, \ldots, T$, a natural strategy is to derive the full conditional distribution of $(a_i^{P1}|a_i^{P2} = 0, a_0, B)$ for $i = 1, \ldots, T$ and $(a_i|a_0, B)$ for $i = T+1, \ldots, N$. Although this is technically feasible, we argue that it involves a lot of matrix computations, and the sampling on a_i^{P1} is computationally intensive because its full conditional distribution is a truncated multivariate normal distribution.

To update the parameters without performing such burdensome computations, we propose an alternative MCMC sampling algorithm here. We have shown that there exists one and only one solution $\Psi = (A, B, a_0, b_0, \Theta, \Sigma)$ to the model defined by Equations (4.1), (4.9) and (4.10) such that A fits the identifiable constraint. Let's call this solution *the identified solution*. Assuming that there is another solution $\tilde{\Psi} = (\tilde{A}, \tilde{B}, \tilde{a}_0, \tilde{b}_0, \tilde{\Theta}, \tilde{\Sigma})$ which does not satisfy the identifiable constraint. We note that it will be easier to generate the random deviates of each parameter in $\tilde{\Psi}$ since there is no constraint on it. Then, if we can find a path to convert $\tilde{\Psi}$ into Ψ, we can obtain the identified solution efficiently. Fortunately, this is achievable through a QR decomposition. Since (A, B) and $(\tilde{A}, \tilde{B})$ are both solutions to Equation (4.1), we have $A'B = \tilde{A}'\tilde{B}$. Since we set the covariance matrix of a_i to be proportional to an identity matrix, there exists an orthogonal matrix Γ such that $\tilde{A} = \Gamma A$ and $\tilde{B} = (\Gamma')^{-1}B$. Noting that Γ is orthogonal and A is upper triangular, $\tilde{A} = \Gamma A$ is like doing a QR decomposition on $\tilde{A}$, and we know that when the diagonal elements of A are restricted to be positive, such decomposition is unique. Therefore, when observing a non-identified $\tilde{A}$, we can easily convert it to the identified A by doing a QR decomposition.

Next, we show that for any two non-identified solutions, $\tilde{\Psi}_1 = (\tilde{A}_1, \tilde{B}_1)$ and $\tilde{\Psi}_2 = (\tilde{A}_2, \tilde{B}_2)$, applying the QR decomposition on each of them will result in the same identified (A, B). Suppose that $\tilde{A}_1 = \Gamma_1 A_1$ and $\tilde{A}_2 = \Gamma_2 A_2$, where Γ_1 and Γ_2 are orthogonal and A_1 and A_2 follow the identifiable constraint. Since $\tilde{\Psi}_1$ and $\tilde{\Psi}_2$ are both solutions to the model, there exists an orthogonal matrix Γ such that $\tilde{A}_1 = \Gamma \tilde{A}_2$; then $\Gamma_1 A_1 = \tilde{A}_1 = \Gamma(\Gamma_2 A_2) = (\Gamma\Gamma_2)A_2 = \tilde{\Gamma} A_2$, where $\tilde{\Gamma} = \Gamma\Gamma_2$ is also an orthogonal matrix. According to the uniqueness of QR decomposition, we will have $\Gamma_1 = \tilde{\Gamma}$ and $A_1 = A_2$.

Since we include a model selectiony procedure in our proposed method, we need to show that the post-processing procedure does not have any effect on the model indicator, γ. In other word, for any two solutions $(B_1, \Theta_1, \Sigma_1)$ and $(B_2, \Theta_2, \Sigma_2)$, which are identical subject to an orthogonal transformation, we should show that $\pi(\gamma|B_1, w) = \pi(\gamma|B_2, w)$. If $B_1 = \Gamma B_2$, we have $\Theta_1 = \Gamma\Theta_2$ and $\Sigma_1 = \Gamma\Sigma_2\Gamma^{'}$. Assuming that the priors on Θ_2 and Σ_2^{-1} are defined as Equations (4.13) and (4.8), then:

$$\Theta_1 = \Gamma\Theta_2 \sim \mathrm{MN}(0, \Sigma_1, H_{(\gamma)}), \tag{4.17}$$

$$\Sigma_1^{-1} = \Gamma\Sigma_2^{-1}\Gamma^{'} \sim W(\nu, \Gamma V \Gamma^{'}). \tag{4.18}$$

So it is easy to observe that $H_{(\gamma)}$, $K_{(\gamma)}$ and $\pi(\gamma|w)$ in Equation (4.16) are the same under B_1 and B_2. Also,

$$\begin{aligned} |V_{(\gamma)}|\Big| B_1 &= |\Gamma V^{-1}\Gamma + B_1 B_1^{'} - B_1 X_{(\gamma)}^{'} K_{(\gamma)}^{-1} X_{(\gamma)} B_1^{'}| \\ &= |\Gamma V^{-1}\Gamma + \Gamma B_2 B_2^{'}\Gamma^{'} - \Gamma B_2 X_{(\gamma)}^{'} K_{(\gamma)}^{-1} X_{(\gamma)} B_2^{'}\Gamma^{'}| \end{aligned}$$

$$= |\Gamma||V^{-1} + B_2 B_2^{'} - B_2 X_{(\gamma)}^{'} K_{(\gamma)}^{-1} X_{(\gamma)} B_2^{'}||\Gamma^{'}|$$

$$= |V^{-1} + B_2 B_2^{'} - B_2 X_{(\gamma)}^{'} K_{(\gamma)}^{-1} X_{(\gamma)} B_2^{'}| = |V_{(\gamma)}|\Big|B_2.$$

Therefore, we can conclude that $\pi(\gamma|B_1, w) = \pi(\gamma|B_2, w)$.

Based on the above argument, we can be assured that it is legitimate to acquire the identified solution by post-processing the non-identified parameters, which can be easily generated from their full conditional distributions through Gibbs sampling. Furthermore, the post-processing action will not affect the variable selection result. The sampling algorithm is illustrated as following:

1. Generate σ^{-2} from $\pi(\sigma^{-2}|\tilde{A}, \tilde{B})$;

2. Generate the non-identified parameters $\tilde{A}, \tilde{a_0}, \tilde{B}, \tilde{\Theta}, \tilde{\Sigma}$ iteratively from their full conditional distribution;

3. Generate γ from $\pi(\gamma|\tilde{B}, w)$ and w from $\pi(w|\gamma)$;

4. Compute the identified parameters: do a QR decomposition on $\tilde{A}$ such that $\tilde{A} = \Gamma A$, where Γ is an orthogonal matrix and A is a matrix satisfying the identifiable constraint; then calculate $a_0 = \Gamma^{'}\tilde{a_0}$, $B = \Gamma^{'}\tilde{B}$,$\Theta = \Gamma^{'}\tilde{\Theta}$ and $\Sigma = \Gamma^{'}\tilde{\Sigma}\Gamma$;

5. Repeat Step (1) – (4).

Note that σ^{-2}, γ and w in Steps (1) and (3) are identified, which do not need post-processing. These steps are repeated iteratively and recursively until converged draws for the identified parameters can be obtained. Finally, the identified parameters from Steps (1), (3) and (4) are recorded and reported.

4.3.5 Full Conditional Distributions

In summary, the proposed Bayesian vector MDS with variable selection procedure assumes:

$$Z = A^{'}B + \Phi, \quad \Phi \sim \sigma \text{MN}(0, I_N, I_J),$$

where A satisfies the identifiable constraint. For any non-identified solution $(\tilde{A}, \tilde{B})$, we adopt the following priors:

$$\begin{aligned}
&\tilde{A}|\tilde{a}_0 \sim \text{MN}(\tilde{A}_0, cI_T, I_N), \\
&\tilde{a}_0 \sim N(0, G_a), \\
&\tilde{B}|\tilde{b}_0, \tilde{\Theta}, \tilde{\Sigma} \sim \text{MN}(\tilde{B}_0 + \tilde{\Theta}X, \tilde{\Sigma}, I_J), \\
&\tilde{b}_0 \sim N(0, G_b), \\
&\tilde{\Theta}_{(\gamma)}|\tilde{\Sigma}, \gamma \sim \text{MN}(0, \tilde{\Sigma}, H_{(\gamma)}), \\
&\tilde{\Sigma} \sim W(\nu, V).
\end{aligned}$$

In addition, priors on the identified parameters are:

$$\begin{aligned}
&\sigma^{-2} \sim \text{Ga}(\frac{m}{2}, \frac{m}{2}), \\
&\gamma_k|w \sim \text{Ber}(w), \\
&w \sim \text{Beta}(p, q).
\end{aligned}$$

The full conditional distributions of the identified/non-identified parameters that will be used in Steps (1) to (3) in the previous section are shown below. The derivation of these full conditional distributions is shown in the Appendix.

- $\sigma^{-2}|$ all others $\sim \text{Ga}(m_1, m_2)$, where:

$$m_1 = \frac{1}{2}(NJ + m) \text{ and } m_2 = \frac{1}{2}\left(m + \text{tr}[(Z - A'B)(Z - A'B)']\right);$$

- $\tilde{A}|$ all others $\sim \text{MN}(\bar{A}, A_l, I_N)$, where:

$$A_l = (\sigma^{-2}\tilde{B}\tilde{B}' + I_T/c)^{-1} \text{ and } \bar{A} = A_l(\sigma^{-2}\tilde{B}Z' + \tilde{A}_0/c);$$

- $\tilde{a}_0|$ all others $\sim N(\bar{a}, G_{an})$, where:

$$G_{an} = \left(NI_T/c + G_a^{-1}\right)^{-1} \text{ and } \bar{a} = G_{an}\left(\sum_{i=1}^{N}\tilde{a}_i\right)/c;$$

- $\tilde{B}|$ all others $\sim \text{MN}(\bar{B}, B_l, I_J)$, where:

$$B_l = (\sigma^{-2}\tilde{A}\tilde{A}' + \tilde{\Sigma}^{-1})^{-1} \text{ and } \bar{B} = B_l(\sigma^{-2}\tilde{A}Z + \tilde{\Sigma}^{-1}(\tilde{B}_0 + \tilde{\Theta}X));$$

- $\tilde{\Theta}_{(\gamma)}|$ all others $\sim \text{MN}(\bar{\Theta}_{(\gamma)}, \tilde{\Sigma}, H_c)$, where:

$$H_c = \left(X_{(\gamma)}X'_{(\gamma)} + H^{-1}_{(\gamma)}\right)^{-1} \text{ and } \bar{\Theta}_{(\zeta)} = \tilde{B}X'_{(\gamma)}H_c;$$

- $\tilde{b}_0|$ all others $\sim N(\bar{b}, G_{bn})$, where:

$$G_{bn} = \left(J\tilde{\Sigma}^{-1} + G_b^{-1}\right)^{-1} \text{ and } \bar{b} = G_{bn}\tilde{\Sigma}^{-1}\sum_{j=1}^{J}\left(\tilde{b}_j - \tilde{\Theta}X_j\right);$$

- $\tilde{\Sigma}^{-1}|$ all others $\sim W(J + \sum_k \gamma_k + \nu, V_n)$, where:

$$V_n = \left(V^{-1} + (\tilde{B} - \tilde{B}_0 - \tilde{\Theta}X)(\tilde{B} - \tilde{B}_0 - \tilde{\Theta}X)' + \tilde{\Theta}_{(\gamma)} H_{(\zeta)}^{-1} \tilde{\Theta}'_{(\gamma)}\right)^{-1};$$

- $w|$ all others $\sim$ Beta$(p + \sum_{k=1}^{K} \gamma_k, q + K - \sum_{k=1}^{K} \gamma_k)$;

- The conditional posterior distribution of $\gamma|\tilde{B}$ is shown in Equation (4.16).

4.4 Simulated Data Example

In this section, we apply our proposed Bayesian vector MDS model with variable selection procedure to two simulated examples. For both examples, we have $N = 100$ consumers, $J = 14$ brands, and the true dimensionality is $T = 4$. In the first example, there are $K = 8$ brand attributes and in the second $K = 14$. We generate the identified σ^2 and the non-identified $\tilde{a}_0$, $\tilde{b}_0$, $\tilde{\Sigma}$ according to their prior distributions, respectively. The odd numbers of γ are set to be 1, which indicates that the corresponding attributes are significant, and the even numbers of γ are 0. The columns of $\tilde{\Theta}$ with odd numbers, $\tilde{\Theta}_{(\gamma)}$, is then generated based on $\tilde{\Sigma}$ and γ according to Equation (4.13), and the columns of $\tilde{\Theta}$ with even number are all zero. After obtaining $\tilde{A}$ and $\tilde{B}$ based on their priors as shown in Equations (4.2) and (4.10), we apply a QR decomposition on $\tilde{A}$ to get the identified parameter A, followed by a transformation on all the other non-identified parameters as demonstrated in Step (4) in Section 4.3.4. Now that we get all identified parameters, especially A and B, the observed consumer rating data Z can be generated according to Equation (4.1).

We compare our model with the classical vector model MDPREF. In addition, to examine the value of the variable selection procedure, we include a second benchmark model, which is similar to our proposed model but there is no variable selection procedure involved. For simplicity, we will call this benchmark model BVMDS, and our proposed model BVMDS-VS. We will compare whether the models are able to identify the true dimensionality, how well they estimate the parameters and whether the proposed model can be used to investigate the significance of each brand attribute. Furthermore, in each data set, there are two validation brands so that we can evaluate each model's prediction accuracy.

4.4.1 Simulated Data Example with $K = 8$

Knowing that the true dimensionality is $T = 4$, we check if the three comparing models can identify the dimension correctly. For the two Bayesian models, BVMDS-VS and BVMDS, we compute the Deviance Information Criterion (DIC) of each model under each dimensionality, and for MDPREF, the variance-accounted-for (VAF) statistic under each dimensionality is used to determine the optimal dimension. Table 4.1 reports the values of these measurements. Both BVMDS-VS and BVMDS show that DIC of the 4-dimensional solution is the smallest, which indicates that they successfully identify the correct dimension, while MDPREF shows that $T = 2$ or $T = 3$ is the optimal dimension. Furthermore, if we compare DIC values of BVMDS-VS and BVMDS for the 4-dimensional solution, we can see that the one with variable selection procedure has a smaller DIC value, indicating that it is a better model over the benchmark.

To compare model's parameter estimation performance, we compute the root mean square error (RMSE) of the estimated parameters. Since we impose the identifiable constraint on BVMDS-VS and BVMDS, the model parameters are fully identified, so we can simply use the posterior mean of each parameter as its estimated value. However, MDPREF is under-identified, and it suffers from all the identifica-

Table 4.1. Simulation Analysis - variable selection Heuristics for the Three Models when the number of brand attributes is smaller than the number of brands.

Model	Criterion	Dim=1	Dim=2	Dim=3	**Dim=4**	Dim=5
BVMDS-VS	DIC	6518.76	6428.14	6418.03	**6386.66**	6390.09
BVMDS	DIC	6462.92	6428.59	6434.54	**6406.10**	6420.13
MDPREF	VAF	0.69	**0.86**	**0.91**	0.92	0.93

Note: **bold** highlights the optimal dimension under each model.

tion issues discussed in Section 4.3.2. As discussed, the classical approaches usually use a Procrustes transformation to address this problem. In Table 4.2, we list the RMSE of BVMDS-VS, BVMDS and MDPREF. It is clearly shown that BVMDS-VS generates the smallest errors for almost all the parameters, and the advantage of the two Bayesian models over the classical approach is also obvious.

Table 4.2. Simulation Analysis - Parameter Estimation Comparison between the Three Models when the number of brand attributes is smaller than the number of brands.

Parameter	BVMDS-VS	BVMDS	MDPREF
σ	0.034	0.032	0.111
A	**0.622**	0.634	0.640
a_0	**0.216**	0.240	0.429
B	**0.279**	0.284	0.337
b_0	**0.164**	0.170	0.198
Θ	**0.090**	0.093	0.094
Σ	**0.001**	0.001	0.007

Note: **bold** indicates that the proposed BVMDS-VS model beats two benchmark models for the parameter of consideration.

Table 4.3 shows the comparative results on the out-of-sample prediction of three competing models. RMSE between the predicted rating of the validation brands and the true rating data is calculated. Note that since MDPREF does not provide explicit information on the optimal dimensionality, we report RMSE based on all $2/3/4-$dimensional solutions. The numbers support the use of our proposed BVMDS-

VS model because it makes the best prediction.

Table 4.3. Simulation Analysis - Out-of-Sample Prediction Comparison between the Three Models when the number of brand attributes is smaller than the number of brands.

Brand	BVMDS-VS 4-Dim Sol.	BVMDS 4-Dim Sol.	MDPREF 2-Dim Sol.	MDPREF 3-Dim Sol.	MDPREF 4-Dim Sol.
Brand1	**0.92**	0.96	1.03	1.01	0.98
Brand2	**0.95**	1.01	1.10	1.01	1.10

Note: **bold** highlights the minimum RMSE among the comparing models with various dimensions when predicting the ratings of the validation brands.

As an important feature of our proposed BVMDS-VS model, we claim that it can identify the significant brand attributes efficiently; meanwhile, during the sampling procedure, only the preferred attributes will be involved to estimate the other parameters. Our model achieves this goal by introducing a latent indicator vector γ, and the final optimal model is the model corresponding to the mode of the posterior distribution of γ. Somehow, the two benchmark models can also provide justification on the significance of each component. For BVMDS, a common practice is to compute the probability of each component to be positive (PP) (Rossi, McCulloch, & Allenby, 1996). At a 0.1 threshold, if PP is greater than 0.9, the corresponding component is labeled as positively significant, and if PP is smaller than 0.1, it is negatively significant. For the classical vector model MDPREF, one can regress the estimated brand vectors (after Procrustes transformation) on the brand attributes (also including an intercept) for each dimension when $K < J$. Then the significance can be calculated by checking the p-value of each coefficient. Table 4.4 presents these measurements. The column of BVMDS-VS lists the indicator vector of the optimal model, which is exactly the same as the true model. The columns of BVMDS shows PP of each Θ component. We observe that although it is doing a good job in identifying the non-significant attributes, it misses several significant components. Attention should be paid that in this particular example, BVMDS fails to label significance on all 4 significant components in Dimension 3. The columns of MDPREF includes the p-value of each Θ component. Similar to BVMDS, it succeeds in all non-significant attributes, but misses some significant components.

Table 4.4. Simulation Analysis - Variable Selection Results of the Three Models when the number of brand attributes is smaller than the number of brands.

Attribute	BVMDS-VS	BVMDS				MDPREF			
		Dim1	Dim2	Dim3	Dim4	Dim1	Dim2	Dim3	Dim4
Attr 1	<u>1</u>	**0.95**	**0.03**	0.86	0.18	**0.00**	0.14	**0.01**	**0.01**
Attr 2	0	0.83	0.32	0.32	0.72	0.80	0.70	0.58	0.63
Attr 3	<u>1</u>	**0.07**	**0.02**	0.37	**0.94**	**0.02**	**0.02**	0.16	**0.00**
Attr4	0	0.27	0.30	0.27	0.71	0.34	0.48	0.90	0.27
Attr 5	<u>1</u>	**0.04**	**0.04**	0.74	**0.93**	0.25	**0.09**	**0.00**	**0.00**
Attr 6	0	0.83	0.61	0.71	0.40	0.43	0.64	0.88	0.26
Attr 7	<u>1</u>	**0.99**	**0.09**	0.36	**0.09**	**0.05**	**0.00**	0.30	**0.00**
Attr 8	0	0.10	0.29	0.81	0.74	1.00	0.99	0.15	0.26

Note:
1. *Italic* attributes are the real significant attributes;
2. <u>1</u> means that the attribute is included in the optimal model based on BVMDS-VS;
3. **Bold** numbers represents that this component is significant based on BVMDS or MDPREF.

4.4.2 Simulated Data Example with $K = 14$

We then consider the performance of the three competing models when the number of attributes is not smaller than the number of brands. In this case, BVMDS actually imposes a prior on B in a regression setting with more regressors than the number of observations, so it is unidentified. Intuitively, such indeterminacy problem will cause trouble in the estimation of Θ. Also, for MDPREF, a stepwise regression is applied on B to obtain an estimate of Θ, which will badly decrease the performance of parameter estimation and out-of-sample prediction accuracy of MDPREF.

We start with checking the power of the three competing models to identify the true underlying dimensionality. Table 4.5 shows that the proposed BVMDS-VS model identifies the correct dimension. BVMDS, however, favors increasing dimensionality, and MDPREF, like the situation when $K = 8$, underestimates the dimension. In addition, by comparing the DIC values of BVMDS-VS and BVMDS when $T = 4$, we find that BVMDS-VS is a better model over BVMDS.

Table 4.5. Simulation Analysis - variable selection Heuristics for the Three Models when the number of brand attributes is greater than or equal to the number of brands.

Model	Criterion	Dim=1	Dim=2	Dim=3	Dim=4	Dim=5
BVMDS-VS	DIC	9959.68	8432.63	7616.91	**6816.54**	6825.62
BVMDS	DIC	9890.70	8442.35	7625.36	6826.81	6744.79
MDPREF	VAF	0.71	**0.85**	**0.90**	0.93	0.95

Note: **bold** highlights the optimal dimension under each model.

Table 4.6 presents the RMSE of the estimated parameters when comparing with the true parameter values. Note again that we conduct Procrustes transformation on MDPREF to ensure that it does not suffer from identification problems. The numbers clearly support BVMDS-VS because its error is the smallest among the three competing models for almost all parameters of interest.

Table 4.6. Simulation Analysis - Parameter Estimation Comparison between the Three Models when the number of brand attributes is greater than or equal to the number of brands.

Variable	BVMDS-VS	BVMDS	MDPREF
σ	**0.020**	0.021	0.190
A	**0.557**	0.565	0.637
a_0	**0.199**	0.204	0.215
B	**0.508**	0.533	0.653
b_0	**0.399**	0.420	1.000
Θ	**0.087**	0.096	0.140
Σ	0.001	0.001	0.004

Note: **bold** indicates that the proposed BVMDS-VS model beats two benchmark models for the parameter of consideration.

Next, we compare the out-of-sample prediction accuracy of BVMDS-VS, BVMDS and MDPREF. Like Table 4.3, we list the results based on the 4-dimensional solutions of the two Bayesian models and $2/3/4-$dimensional solutions of MDPREF. As expected, the RMSE between the prediction from the models and the true rating data shows that our model outperforms the two benchmark models. Especially notable is its advantage over the classical approach; the RMSEs for both two validation brands from BVMDS-VS are just a half of those from MDPREF.

Table 4.7. Simulation Analysis - Out-of-Sample Prediction Comparison between the Three Models when the number of brand attributes is greater than or equal to the number of brands.

Brand	BVMDS-VS	BVMDS	MDPREF		
	4-Dim Sol.	4-Dim Sol.	2-Dim Sol.	3-Dim Sol.	4-Dim Sol.
Brand1	**1.36**	2.00	2.82	2.78	2.78
Brand2	**1.42**	2.04	3.09	2.85	2.81

Note: **bold** highlights the minimum RMSE among the comparing models with various dimensions when predicting the ratings of the validation brands.

Lastly, we present the variable selection output in Table 4.8. Still, we use 0.1 as the threshold to evaluate significance. First, the optimal model based on BVMDS-VS procedure matches the true model perfectly. For BVMDS and MDPREF, it mistakenly marks significance for some attributes. For example, Attribute 12 is tagged as a contributing variable by both benchmark models. In addition, like the situation when $K = 8$, both of these benchmark models miss some significant components.

Table 4.8. Simulation Analysis - Variable Selection Results of the Three Models when the number of brand attributes is greater than or equal to the number of brands.

Attribute	BVMDS-VS	BVMDS				MDPREF			
		Dim1	Dim2	Dim3	Dim4	Dim1	Dim2	Dim3	Dim4
Attr 1	1	**0.05**	**0.94**	**0.99**	**0.93**	**0.00**	**0.00**	**0.00**	**0.01**
Attr 2	0	0.67	0.80	0.29	0.76	0.25	**0.03**	**0.09**	0.49
Attr 3	1	0.86	**0.99**	0.35	0.32	0.47	**0.00**	**0.00**	**0.00**
Attr 4	0	0.57	0.77	0.39	0.21	0.93	0.11	0.97	**0.06**
Attr 5	1	**0.90**	**0.06**	**0.95**	**0.06**	**0.00**	**0.00**	**0.08**	0.20
Attr 6	0	0.59	0.25	0.71	0.44	0.14	0.56	0.94	0.98
Attr 7	1	**0.02**	**0.01**	0.45	**0.94**	**0.00**	**0.00**	**0.00**	**0.00**
Attr 8	0	0.25	0.65	0.36	0.44	0.52	0.18	0.44	0.80
Attr 9	1	**0.03**	**0.05**	**0.94**	**0.95**	0.84	0.38	**0.00**	**0.00**
Attr 10	0	0.24	0.56	0.48	0.49	0.79	0.87	0.72	0.46
Attr 11	1	**0.93**	**0.02**	0.59	**0.02**	**0.00**	**0.00**	0.98	**0.03**
Attr 12	0	0.52	**0.93**	**0.08**	0.69	**0.08**	0.37	0.19	0.49
Attr 13	1	0.70	**0.98**	0.87	**0.00**	**0.00**	0.72	0.54	**0.00**
Attr 14	0	0.59	0.70	0.48	0.40	0.14	0.51	0.95	0.88

Note:
1. *Italic* attributes are the real significant attributes;
2. 1 means that the attribute is included in the optimal model based on BVMDS-VS;
3. **Bold** numbers represents that this component is significant based on BVMDS or MDPREF.

4.5 Application: SUV Rating Data

4.5.1 The Large SUV Data Set

A large US automotive consumer research supplier administered a tracking study to gauge automotive marketing awareness and shopping behavior among typical consumers: 200 to 300 respondents were asked to provide their image attribute ratings in 16 car segments, and ten light truck and sport utility vehicle (SUV) segments. Typically, "SUV" means a passenger vehicle with off-road and towing capabilities. Sometimes, people treat SUVs like a smaller-sized pickup truck. As the SUV market can be commonly categorized into compact SUVs, large SUVs and luxury SUVs, the large SUVs is the main focus in this application study. The data were collected in December 2002 with 279 consumer intenders rating 16 different large SUVs. By "indender", we mean the consumers who plan to or will purchase a vehicle in the next $6-12$ months. We divide the 16 vehicles included in this image attribute rating data into 14 calibration brands and 2 validation brands. These 14 calibration brands account for 97.6% of 2002 large SUV sales. The resulting 14 brands for the analysis are: Chevy Suburban, Chevy Tahoe, Ford Expedition, Ford Excursion, Lincoln Navigator, Toyota Sequoia, Toyota Land Cruiser, GMC Yukon, Escalade ESV, Escalade EXT, Hummer H1, Lexus LX470, Yukon Denali. The two brands used for prediction/validation are: Mercedes G-Class and Hummer H2. Among all these brands, 18.75% are Ford brands and 68.75% are US brands.

Furthermore, we have access to a data set that include 21 subjectively rated attributes, and that covers measurements on the physical, economical and reliability characteristics of a vehicle. The attribute names and their descriptive statistics are shown in Table 4.9. In addition to these attributes, we also create two dummy variables for "Ford vs. not Ford" and "Domestic vs. Foreign" to be included in the analysis. After creating the joint space presentation of the consumer and brand locations based on the rating data, we wish to relate the resulting latent dimensions to some attributes to study the managerial implications.

4.5.2 The Proposed BVMDS-VS Model Solution

Table 4.10 presents the DIC measurements for the proposed BVMDS-VS method by dimension. As shown, the four-dimensional solution appears to be the optimal.

We show the resulting joint space under the four-dimensional solution in Figure 4.1, where the brand coordinates and the consumer vectors are plotted in pair-wise dimension fashion. The consumer vectors[1] are represented as black dots, and brand

[1] The consumer vectors should be connected to the origin, and point in the direction of the

Table 4.9. Empirical Analysis - Descriptive Statistics of Brand Attributes.

Attribute	Mean	Std
Market share	0.06	0.08
Gas mileage	4.60	0.42
Value for money	5.72	0.50
Workmanship	6.85	0.26
Good ride handling	6.82	0.61
Luxurious	6.73	0.94
Safety	7.16	0.26
Rugged	6.75	0.68
Towing capacity	6.84	0.70
Low price	5.10	0.68
Sporty	6.16	0.21
Good looking	6.68	0.62
Good for family usage	6.97	0.88
Fun to drive	6.88	0.27
Easy to enter/exit	6.91	0.47
Dependable	7.09	0.26
Good acceleration	6.53	0.35
Big cargo space	6.93	0.69
Lasts a long time	6.99	0.25
Prestigious	7.00	0.66
Good trade-in value	6.64	0.33
Ford vs. not Ford	0.19	0.40
Domestic vs. Foreign	0.69	0.48

Table 4.10. Empirical Analysis - DIC of BVMDS-VS under various dimensionality.

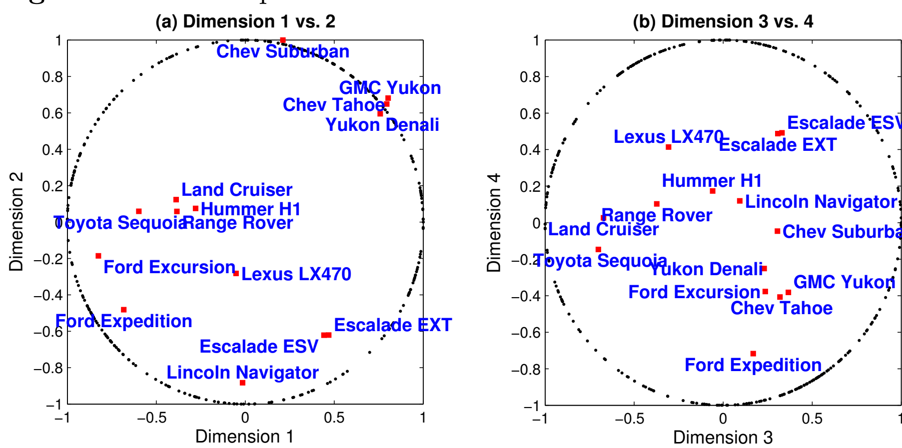
Model	Dim=2	Dim=3	Dim=4	Dim=5
BVMDS-VS	9142.05	8713.59	**7517.79**	7569.12

coordinates as red squares with appropriate names next to them. Note that consumer vectors are normalized to equal length for ease of interpretation. The vector model assumes the ratings are approximated to the inner product of consumer vector and brand coordinate, which measures the projection of the consumer vector to the brand vector. For each brand vector, it is generally true that the projection is larger if

consumer utility

the end of one consumer vector is close to the end of the brand vector. Therefore, such plots provide straightforward presentation on the consumer preference. More specifically, a consumer will generally give a higher rating to the brand projecting further out in the direction of increasing utility. So by looking at the figure, we can tell the preferred consumers or potential customer intenders for each brand.

Figure 4.1. Joint Space of Four-dimensional solution from BVMDS-VS.



Before investigating the interpretation underlying each dimension, we consider an important feature of the proposed model: variable selection. As discussed in Section 4.3.3, the optimal model is determined based on the posterior distribution of the latent indicator vector γ. Table 4.11 presents the indicator vectors with the top 5 posterior probability. The components of the latent vector match the brand attributes listed in Table 4.9 in order. The last two components refer to "Ford" dummy variable and "Domestic" dummy variable. The top model based on our proposed BVMDS-VS indicates that 9 attributes are useful in the estimation of brand coordinates, and this model has a 16.74% chance of selection during the MCMC procedure.

Since we now know that the model with 9 attributes is optimal, we use these attributes to ascribe interpretations to our derived four dimensions. Table 4.12 presents the correlations between the brand coordinates and the attributes involved in the opti-

Table 4.11. Empirical Analysis - variable selection Result of BVMDS-VS.

Variable Name	Model 1	Model 2	Model 3	Model 4	Model 5
Market share	1	1	1	1	1
Gas mileage	1	1	1	1	1
Value for money	1	1	1	1	1
Workmanship	0	0	0	0	0
Good ride handling	0	0	0	0	1
Luxurious	1	0	1	1	0
Safety	0	0	0	0	0
Rugged	0	0	1	1	1
Towing capacity	0	0	1	1	1
Low price	1	1	1	1	1
Sporty	0	0	0	0	0
Good looking	1	1	1	1	1
Good for family usage	0	1	1	1	0
Fun to drive	0	0	0	0	0
Easy to enter/exit	0	0	0	0	0
Dependable	0	0	0	0	0
Good acceleration	0	0	0	0	0
Big cargo space	1	1	1	0	0
Lasts a long time	0	0	0	0	0
Prestigious	0	0	0	0	0
Good trade-in value	0	0	0	0	0
Ford vs. not Ford	1	1	0	0	0
Domestic vs. Foreign	1	1	0	0	0
Model Probability	16.74%	7.71%	3.75%	3.46%	3.37%

mal model. As shown, dimension 1 describes an *aesthetic* latent construct having high correlation with "good looking" (0.62). Dimension 2 appears to be related to a *luxury* dimension having relatively large negative correlations with "luxurious" (−0.59). Dimension 3 is reflecting a *practicality* latent structure that is highly related to domestic makes (0.92) and "big cargo space" (0.79). Finally, dimension 4 is definitely characterized by *price* as it is related to high price (0.76).

Managerially, the joint space maps in Figure 4.1 also provide interesting information to the users, especially regarding the positioning of different brands from the same manufacturer. The most obvious phenomenon a occurs between the two Cadillac

Table 4.12. Empirical Analysis - Correlations between Brand Attributes and Dimensions.

Attribute	Dimension 1	Dimension 2	Dimension 3	Dimension 4
Market share	0.27	0.45	0.41	-0.65
Gas mileage	0.23	0.00	-0.40	0.16
Value for money	0.46	0.56	0.26	-0.53
Luxurious	0.36	-0.59	0.24	0.59
Low price	0.33	0.59	0.34	-0.76
Good looking	0.62	-0.15	0.49	0.05
Big cargo space	0.51	0.36	0.79	-0.41
Ford vs not Ford	-0.50	-0.51	0.21	-0.43
Domestic vs Foreign	0.43	0.03	0.92	-0.24

Escalade brands (Escalade ESV and Escalade EXT). Their brand coordinates in all dimensions are nearly overlapping, which indicates that they are attracting the same group of consumer intenders so that they are basically competing with one another. Since the number of intenders in a specific group remains constant, this simply indicates that these two brother brands are cannibalizing each other's market share. We observe the similar problem on a group of GM brands including GMC Yukon, Yukon Denali, and Chevy Tahoe. The locations of two Toyota brands, Sequoia and Land Cruiser, are also close but a little bit better than the first two. Three Ford sponsored brands are doing a good job in positioning themselves differently from each other, which avoids competition among the same manufacturer. In addition, Hummer H1 and Lexus LX470 have also made themselves distinct from the brands with the same manufacturer, making them less vulnerable to competitive actions.

4.5.3 Comparison with the Benchmark Models

In addition to the proposed procedure, we also run BVMDS and MDPREF on this same data set for comparison purposes. Table 4.13 presents the DIC measurements for the BVMDS-VS and BVMDS model by dimension, while Table 4.14 lists the VAF statistic for MDPREF for dimensions up to 14. The DIC measurement of BVMDS favors increasing dimensionality, which is similar to the simulation result when the number of attributes is more than the number of brands. On the other hand, the marginal improvements on VAF of MDPREF when $T = 5, 6, 7$ are close, so it is hard for us to declare the optimal dimension. Additionally, if we compare the DIC values of BVMDS-VS and BVMDS when $T = 4$, we can see that BVMDS-VS beats BVMDS.

Table 4.13. Empirical Analysis - DIC of BVMDS-VS and BVMDS under various dimensionality.

Model	Dim=2	Dim=3	Dim=4	Dim=5
BVMDS-VS	9142.05	8713.59	**7517.79**	7569.12
BVMDS	9188.53	8664.25	8050.20	7319.62

Table 4.14. Empirical Analysis - VAF of MDPREF under various dimensionality.

Dim	VAF
1	0.26
2	0.42
3	0.55
4	0.64
5	0.72
6	0.78
7	0.83
8	0.87
9	0.91
10	0.94
11	0.97
12	0.99
13	1.00
14	1.00

Furthermore, we make use of the two validation brands to compare the prediction capability of the three models. Similar to the simulation study, the RMSE between the predicted rating and the truly observed rating data is presented in Table 4.15.

Without any surprise, BVMDS-VS is the best model amongst the three competing options when considering both brands. BVMDS does not perform too badly, while the advantage over MDPREF is obvious.

Table 4.15. Empirical Analysis - Out-of-Sample Prediction Comparison between the Three Models when the number of brand attributes is greater than or equal to the number of brands.

Brand	BVMDS-VS	BVMDS	MDPREF		
	4-Dim Sol.	4-Dim Sol.	4-Dim Sol.	5-Dim Sol.	6-Dim Sol.
Mercedes G-Class	**0.58**	0.61	0.73	0.73	0.73
Hummer H2	**0.71**	0.75	0.81	0.81	0.81

Note: **bold** highlights the minimum RMSE among the comparing models with various dimensions when predicting the ratings of the validation brands.

4.6 Conclusion

We introduce a new Bayesian vector MDS with variable selection procedure for the analysis of rating data typically collected in marketing positioning studies. The advantages of the proposed Bayesian method are:

1. The underlying dimensionality can be determined by a well-established criterion;

2. We make use of a related data set to better estimate the brand coordinates. For both simulation examples and the application study, the proposed method shows better performance over the classical procedure, namely MDPREF, which assures the contribution of involving such extra information in the model construction;

3. A variable selection procedure is incorporated in our proposed Bayesian framework to identify the optimal model that contains significant brand attributes.

The simulation and empirical results show that the model with variable selection dominates the one without variable selection procedure, and it is also managerially meaningful;

4. We discuss identification issues of the vector MDS model, and impose proper constraints on the model to solve the indeterminacy problem. In addition, a sampling algorithm is demonstrated to obtain draws of the identified parameters efficiently.

Concerning future research opportunities, there are a number of potential extensions to our model. For example, one may consider modifying this model to analyze ordered preference data by treating them as metric. Also, we only consider a brand-related data set in this study; however, if data are available, one can also involve the consumer-based demographics, psychographics, attitudes, etc. information to construct a similar prior on the consumer vectors, and apply the variable selection procedure. Such model may yield a better positioning strategy and get insights into the heterogeneity of consumer preference. Another generalization of this model is to consider a dynamic Bayesian vector model as the brand positioning and consumer cognition can be changing over time. It will be interesting to check how the brands are moving in the underlying space and how the targeted consumers are changing over time.

Appendix

- The full conditional distribution of σ^{-2} is:

$$\begin{aligned}
&\pi(\sigma^{-2}|\text{ all others})\\
\propto&(\sigma^{-2})^{NJ/2}\text{etr}\bigg\{-\frac{\sigma^{-2}}{2}(Z-A'B)(Z-A'B)'\bigg\}(\sigma^{-2})^{\frac{m}{2}-1}\exp\bigg\{-\frac{m\sigma^{-2}}{2}\bigg\}\\
=&(\sigma^{-2})^{NJ/2+m/2-1}\exp\bigg\{-\sigma^{-2}\bigg[\frac{m}{2}+\frac{1}{2}\text{tr}[(Z-A'B)(Z-A'B)']\bigg]\bigg\}.
\end{aligned}$$

So $\sigma^{-2}|$ all others $\sim \text{Ga}(m_1, m_2)$, where:

$$m_1 = \frac{1}{2}(NJ+m) \text{ and } m_2 = \frac{1}{2}\left(m + \text{tr}[(Z-A'B)(Z-A'B)']\right).$$

- The full conditional distribution of $\tilde{A}$ is:

$$\begin{aligned}
&\pi(\tilde{A}|\text{ all others})\\
\propto&\text{etr}\left\{-\frac{1}{2}\left[\sigma^{-2}(Z-\tilde{A}'\tilde{B})(Z-\tilde{A}'\tilde{B})' + (\tilde{A}-\tilde{A}_0)'(\tilde{A}-\tilde{A}_0)/c\right]\right\}\\
\propto&\text{etr}\left\{-\frac{1}{2}\left[\tilde{A}'(\sigma^{-2}\tilde{B}\tilde{B}'+I_T/c)\tilde{A} - 2\tilde{A}'(\sigma^{-2}\tilde{B}Z'+\tilde{A}_0/c)\right]\right\}.
\end{aligned}$$

So $\tilde{A}|$ all others $\sim \text{MN}(\bar{A}, A_l, I_N)$, where:

$$A_l = (\sigma^{-2}\tilde{B}\tilde{B}' + I_T/c)^{-1} \text{ and } \bar{A} = A_l(\sigma^{-2}\tilde{B}Z' + \tilde{A}_0/c).$$

- The full conditional distribution of $\tilde{a}_0$ is:

$$\begin{aligned}&\pi(\tilde{a}_0|\text{ all others})\\ \propto&\exp\left\{-\frac{1}{2}\sum_{i=1}^{N}\left[(\tilde{a}_i-\tilde{a}_0)'(\tilde{a}_i-\tilde{a}_0)/c\right]-\frac{1}{2}\tilde{a}_0'G_a^{-1}\tilde{a}_0\right\}\\ \propto&\exp\left\{-\frac{1}{2}\left[\tilde{a}_0'\left(NI_T/c+G_a^{-1}\right)\tilde{a}_0-2\tilde{a}_0'/c\left(\sum_{i=1}^{N}\tilde{a}_i\right)\right]\right\}.\end{aligned}$$

So $\tilde{a}_0|$ all others $\sim N(\bar{a}, G_{an})$, where:

$$G_{an}=\left(NI_T/c+G_a^{-1}\right)^{-1}\text{ and }\bar{a}=G_{an}\left(\sum_{i=1}^{N}\tilde{a}_i\right)/c.$$

- The full conditional distribution of $\tilde{B}$ is:

$$\begin{aligned}&\pi(\tilde{B}|\text{ all others})\\ \propto&\text{etr}\left\{-\frac{1}{2}\left[\sigma^{-2}(Z-\tilde{A}'\tilde{B})'(Z-\tilde{A}'\tilde{B})+(\tilde{B}-\tilde{B}_0-\tilde{\Theta}X)'\tilde{\Sigma}^{-1}(\tilde{B}-\tilde{B}_0-\tilde{\Theta}X)\right]\right\}\\ \propto&\text{etr}\left\{-\frac{1}{2}\left[\tilde{B}'(\sigma^{-2}\tilde{A}\tilde{A}'+\tilde{\Sigma}^{-1})\tilde{B}-2\tilde{B}'(\sigma^{-2}\tilde{A}Z+\tilde{\Sigma}^{-1}(\tilde{B}_0+\tilde{\Theta}X))\right]\right\}.\end{aligned}$$

So $\tilde{B}|$ all others $\sim \text{MN}(\bar{B}, B_l, I_J)$, where:

$$B_l=(\sigma^{-2}\tilde{A}\tilde{A}'+\tilde{\Sigma}^{-1})^{-1}\text{ and }\bar{B}=B_l(\sigma^{-2}\tilde{A}Z+\tilde{\Sigma}^{-1}(\tilde{B}_0+\tilde{\Theta}X)).$$

- The full conditional distribution of $\tilde{b}_0$ is:

$$
\begin{aligned}
&\pi(\tilde{b}_0 | \text{ all others}) \\
\propto &\exp\left\{-\frac{1}{2}\sum_{j=1}^{J}\left[(\tilde{b}_j - \tilde{b}_0 - \tilde{\Theta}X_j)'\tilde{\Sigma}^{-1}(\tilde{b}_j - \tilde{b}_0 - \tilde{\Theta}X_j)\right] - \frac{1}{2}\tilde{b}_0' G_b^{-1}\tilde{b}_0\right\} \\
\propto &\exp\left\{-\frac{1}{2}\left[\tilde{b}_0'\left(J\tilde{\Sigma}^{-1} + G_b^{-1}\right)\tilde{b}_0 - 2\tilde{b}_0'\tilde{\Sigma}^{-1}\sum_{j=1}^{J}\left(\tilde{b}_j - \tilde{\Theta}X_j\right)\right]\right\}.
\end{aligned}
$$

So $\tilde{b}_0|$ all others $\sim N(\bar{b}, G_{bn})$, where:

$$
G_{bn} = \left(J\tilde{\Sigma}^{-1} + G_b^{-1}\right)^{-1} \text{ and } \bar{b} = G_{bn}\tilde{\Sigma}^{-1}\sum_{j=1}^{J}\left(\tilde{b}_j - \tilde{\Theta}X_j\right).
$$

- The full conditional distribution of $\tilde{\Theta}_{(\gamma)}$ is:

$$
\begin{aligned}
&\pi(\tilde{\Theta}_{(\gamma)} | \text{ all others}) \\
\propto &\text{etr}\left\{-\frac{1}{2}\left[(\tilde{B} - \tilde{\Theta}_{(\gamma)}X_{(\gamma)})'\tilde{\Sigma}^{-1}(\tilde{B} - \tilde{\Theta}_{(\gamma)}X_{(\gamma)}) + \tilde{\Theta}_{(\gamma)}H_{(\gamma)}^{-1}\tilde{\Theta}_{(\gamma)}'\tilde{\Sigma}^{-1}\right]\right\} \\
\propto &\text{etr}\left\{-\frac{1}{2}\left[\tilde{\Theta}_{(\gamma)}\left(X_{(\zeta)}X_{(\gamma)}' + H_{(\gamma)}^{-1}\right)\tilde{\Theta}_{(\gamma)}'\tilde{\Sigma}^{-1} - 2\tilde{\Sigma}^{-1}BX_{(\zeta)}'\tilde{\Theta}_{(\gamma)}'\right]\right\}.
\end{aligned}
$$

So $\tilde{\Theta}_{(\gamma)}|$ all others $\sim \text{MN}(\bar{\Theta}_{(\gamma)}, \tilde{\Sigma}, H_c)$, where:

$$
H_c = \left(X_{(\gamma)}X_{(\gamma)}' + H_{(\gamma)}^{-1}\right)^{-1} \text{ and } \bar{\Theta}_{(\zeta)} = \tilde{B}X_{(\gamma)}'H_c.
$$

- The full conditional distribution of $\tilde{\Sigma}^{-1}$ is:

$$\begin{aligned}
&\pi(\tilde{\Sigma}^{-1}| \text{ all others}) \\
\propto&\left|\tilde{\Sigma}^{-1}\right|^{\frac{J}{2}} \operatorname{etr}\left\{-\frac{1}{2}(\tilde{B}-\tilde{B}_0-\tilde{\Theta}X)'\tilde{\Sigma}^{-1}(\tilde{B}-\tilde{B}_0-\tilde{\Theta}X)\right\} \\
&\times \left|\tilde{\Sigma}^{-1}\right|^{\frac{d-T+1}{2}} \operatorname{etr}\left\{-\frac{1}{2}\tilde{\Sigma}^{-1}V^{-1}\right\} \left|\tilde{\Sigma}^{-1}\right|^{\frac{\sum_k \gamma_k}{2}} \operatorname{etr}\left\{-\frac{1}{2}\tilde{\Theta}_{(\gamma)}H_{(\gamma)}^{-1}\tilde{\Theta}'_{(\gamma)}\tilde{\Sigma}^{-1}\right\} \\
\propto&\left|\tilde{\Sigma}^{-1}\right|^{\frac{J+\sum_k \gamma_k+\nu-T+1}{2}} \\
&\times \operatorname{etr}\left\{-\frac{1}{2}\tilde{\Sigma}^{-1}\left(V^{-1}+(\tilde{B}-\tilde{B}_0-\tilde{\Theta}X)(\tilde{B}-\tilde{B}_0-\tilde{\Theta}X)'+\tilde{\Theta}_{(\gamma)}H_{(\zeta)}^{-1}\tilde{\Theta}'_{(\gamma)}\right)\right\}.
\end{aligned}$$

So $\tilde{\Sigma}^{-1}|$ all others $\sim W(J+\sum_k \gamma_k+\nu, V_n)$, where:

$$V_n = \left(V^{-1}+(\tilde{B}-\tilde{B}_0-\tilde{\Theta}X)(\tilde{B}-\tilde{B}_0-\tilde{\Theta}X)'+\tilde{\Theta}_{(\gamma)}H_{(\zeta)}^{-1}\tilde{\Theta}'_{(\gamma)}\right)^{-1}.$$

- The full conditional distribution of w is:

$$\begin{aligned}
&\pi(w| \text{ all others}) \\
\propto& w^{p-1}(1-w)^{q-1}\prod_{k=1}^{K}\left(w^{\gamma_k}(1-w)^{1-\gamma_k}\right) \\
\propto& w^{p+\sum_{k=1}^K \gamma_k-1}(1-w)^{q+K-(\sum_{k=1}^K \gamma_k)-1}.
\end{aligned}$$

So $w|$ all others $\sim \text{Beta}(p+\sum_{k=1}^K \gamma_k, q+K-\sum_{k=1}^K \gamma_k)$.

Chapter 5

Discussion and Future Work

This chapter summarizes the work performed in this dissertation and highlights the important results and findings in the previous chapters. In addition, model generalizations and future research directions are discussed.

5.1 Summary of Contributions

In Chapter 2, we utilize a heterogeneous Bayesian dynamic model to investigate the relationship between customer satisfaction and firm financial performance. The model provides firm-level estimates of the association, and helps to identify the effects of customer satisfaction on a firm's financial performance over time. It does not need to separate each part-worth into static firm-level and aggregate time-varying components as in Liechty et al. (2005), so the proposed model is more general and less restrictive. Also, our model allows the use of demographic/firmographic variables to explain the variation in firm heterogeneity and time dependent part-worths which is an advantage over the traditional dynamic linear models. Furthermore, the proposed model includes a time varying impact coefficient matrix to explore the dynamic phenomenon which has not been studied in the literature before.

In Chapter 3, we propose a general Bayesian random-coefficient multinomial probit model to analyze consumer choice panel data. The model uses the maximum allowable number of free variance-covariance parameters such that it is still identifiable. Because only relative utilities matter when doing statistical inference, The resulting model assumption is general. Our model yields unequal variances of the relative brand utilities and unequal covariances between each pair of the relative brand utilities, which allows for both positive and negative values. In addition, the new model is less sensitive to prior specifications compared with the one (RMA model) given in Rossi, et al. (1996). We have shown that the posterior distribution from the RMA model becomes improper in the limit when there is only one observation per household and some prior variances are set larger and larger. In contrast, our model always gives proper posterior distributions. Moreover, we employ a parameter expansion technique to develop an efficient algorithm to implement our methodology.

In Chapter 4, we present a Bayesian vector multidimensional scaling with variable selection procedure to study consumer rating data. The method provides a joint space map of consumers and brands in an optimal dimensionality. We impose appropriate constraints on the model to solve the identification problems, and develop a sampling algorithm to draw the identified parameter efficiently. To compare models, we utilize the DIC statistic which takes account of the different number of free parameters among various models. It is important to note that the proposed model incorporates an effective variable selection procedure to determine the significant variables contributing to the relative brand positions on the joint space map with substantial managerial implication for optimal product positioning. Both simulation and empirical data results show that the proposed model outperforms a benchmark model which does not include the variable selection procedure.

5.2 Future Research

For future research, we list below several possible directions.

For the heterogeneous Bayesian dynamic model in Chapter 2, recall that we use the firmographic variables to explain the variation on firm heterogeneity and time dependent part-worths. Yet, not all the independent variables contribute to the explanation of the variation. The current methodology relies on the probability of each parameter being positive to evaluate the significance of these variables. However, if variable selection is the main focus of the study, it will be preferable to incorporate a variable selection procedure in the model to formally choose variables. A possible approach is to employ the idea discussed in Section 4.3.3. We can assume a latent indicator variable ζ, which is embedded in the prior specification of the impact coefficient matrix of the firmographic variables. Then, assuming a multivariate normal prior on the vectorization of the reduced impact coefficient matrix determined by ζ, ζ can be sampled according to its full conditional distribution as shown in Equation (4.16). After tabulation, ζ with the highest posterior probability will be treated as the indicator of the significance of each variable. If one believes that the significance of the variables may be changing over time, we can introduce a dynamic indicator variable ζ_t instead. However, more work is needed to address this problem.

For the Bayesian random-coefficient multinomial probit model presented in Chapter 3, a generalization is to provide time-varying part-worths in addition to individual heterogeneity. When a panel data set is balanced, that is, the total number of purchases for each household is identical, this extension should be relatively easy to achieve. One may use the methodology proposed in Chapter 2 to model the latent brand utility which involves time-varying regression coefficients. When a panel data set is unbalanced, things get complicated, and more work need be done to develop a model with household-level dynamic coefficients.

For the Bayesian vector multidimensional scaling with model selection procedure discussed in Chapter 4, a straightforward extension is to modify the model to analyze ordered preference data. For example, if there are C response categories, one will first assume $C-1$ latent cutoff points γ such that a consumer's latent utility falls between γ_{c-1} and γ_c when the preference rating is c (see Fong et al. 2010). The latent utility is then assumed to be described by our BVMDS-VS model. When consumer-based information, such as demographics, psychographics, attitudes, etc, are available, it seems natural to incorporate these data into the analysis and one can impose a model selection structure in the model as well. Lastly, it will be of interest to develop a dynamic BVMDS model. As things change over time, it will be revealing to see how the brands and selected consumers move within the underlying space.

References

Aguilar, O., & West, M. (2000). Bayesian dynamic factor models and portfolio allocation. *Journal of Business and Economic Statistics*, *18*, 338–57.

Aksoy, L., Cooil, B., Christophe, G., Keiningham, T. L., & Yalcin, A. (2008). The long-term stock market valuation of customer satisfaction. *Journal of Marketing*, *72*(4), 105–22.

Allenby, G. M., & Ginter, J. L. (1995). Using extremes to design products and segment markets. *Journal of Marketing Research*, *32*(4), 392–403.

Anderson, E. W., Fornell, C., & Lehmann, D. R. (1994). Customer satisfaction, market share, and profitability: Findings from sweden. *Marketing Science*, *58*, 53–66.

Anderson, E. W., Fornell, C., & Mazvancheryl, S. (2004). Customer satisfaction and shareholder value. *Journal of Marketing*, *68*(4), 172–85.

Ansari, A. M., & Mela, C. F. (2003). E-customization. *Journal of Marketing Research*, *40*(5), 131–45.

Benzécri, J. (1992). *Correspondence analysis handbook.* Marcel Dekker.

Bernardo, J. M., & Smith, A. F. M. (1994). *Bayesian theory.* Chichester: Wiley.

Bhaumik, A., Dey, D. K., & Ravishanker, N. (2003). *A dynamic linear model approach for compositional time series analysis* (Tech. Rep.). University of Connecticut.

Blei, D. M., Jordan, M. I., & Ng, A. Y. (2003). Hierarchical bayesian models for applications in information retrieval. *Bayesian Statistics*, *7*, 25–43.

Bodapati, A., & Gupta, S. (2003, Feburary). A direct approach to predicting discretized response in target marketing. *Journal of Marketing Research*, *41*, 73–85.

Borg, I., & Groenen, P. (2005). *Modern Multidimensional Scaling: Theory and*

Applications. Springer.

Broet, P., Richardson, S., & Radvanyi, F. (2002). Bayesian hierarchical model for identifying changes in gene expression from microarray experiments. *Journal of computational biology*, *9*(4), 671–83.

Brown, P. J., Vannucci, M., & Fearn, T. (1998). Multivariate bayesian variable selection and prediction. *Journal of The Royal Statistical Society Series B-statistical Methodology*, *60*, 627–41.

Carroll, J., & Chang, J.-J. (1970). Analysis of individual differences in multidimensional scaling via an n-way generalization of "eckart-young" decomposition. *Psychometrika*, *35*(3), 283-319.

Cattell, R. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, *1*, 245–76.

Chen, Y., Narasimhan, C., & Zhang, Z. J. (2001). Individual marketing with imperfect targetability. *Marketing Science*, *20*(1), 23–41.

Chib, S., & Greenberg, E. (1995). Understanding the metropolis-hastings algorithm. *The American Statistician*, *49*(4), 327–35.

Cowles, M. K. (1996). Accelerating monte carlo markov chain convergence for cumulative-link generalized linear models. *Statistics and Computing*, *6*, 101–11.

Dawid, A. (1981). Some matrix-variate distribution theory: Notational considerations and a bayesian application. *Biometrika*, *68*(1), 265–74.

DeSarbo, W. S., Fong, D. K. H., Liechty, J., & Coupland, J. (2005, March). Evolutionary preference/utility functions: A dynamic perspective. *Psychometrika*, *70*(1), 179–202.

DeSarbo, W. S., Kim, Y., & Fong, D. K. H. (1999). A bayesian multidimensional scaling procedure for the spatial analysis of revealed choice data. *Journal of Econometrics*, *89*, 79–108.

DeSarbo, W. S., Park, J., & Rao, V. (2011). Deriving joint space positioning maps from consumer preference ratings. *Marketing Letters*, *22*(1), 1-14.

Fong, D. K. H., DeSarbo, W. S., Park, J., & Scott, C. J. (2010). A bayesian vector multidimensional scaling procedure for the analysis of ordered preference data. *Journal of the American Statistical Association*, *105*(490), 482-92.

Fong, D. K. H., Ebbes, P., & DeSarbo, W. S. (2011). *A new hierarchical bayesian regression model for cross sectional data: an application in examining the drivers of perceived value.*

Fornell, C., Johnson, M. D., Anderson, E. W., Cha, J., & Bryant, B. E. (1996). The american customer satisfaction index: Nature, purpose, and findings. *Journal of Marketing*, *60*(4), 7–18.

Fornell, C., Mithas, S., Morgeson, F. V. I., & Krishnan, M. S. (2006). Customer satisfaction and stock prices: High returns, low risk. *Journal of Marketing*, *70*(1), 3–14.

Gelfand, A. E., & Smith, A. F. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, *85*(410), 398–409.

George, E. I., & McCulloch, R. E. (1997). Approaches for bayesian variable selection. *Statistica Sinica*, *7*, 339–73.

Ghosh, M., Dutta, S., & Stremersch, S. (2006). Customizing complex products: When should the vendor take control? *Journal of Marketing Research*, *43*, 664–79.

Gifi, A. (1990). *Nonlinear multivariate analysis*. Chichester, England: Wiley.

Gormley, I. C., & Murphy, T. B. (2006). A latent space model for rank data. In *Icml'06 proceedings of the 2006 conference on statistical network analysis.* Springer-Verlag.

Graves, S. C., Kletter, D. B., Hetzel, W. B., & Bolton, R. N. (1998). A dynamic model of the duration of the customer's relationship with a continuous service provider: the role of satisfaction. *Marketing Science*, *17* (1), 45–65.

Gruca, T. S., & Rego, L. L. (2005). Customer satisfaction, cash flow, and shareholder value. *Journal of Marketing*, *69* (3), 115–30.

Hastings, W. K. (1970). Monte carlo sampling methods using markov chains and their applications. *Biometrika*, *57*, 97–109.

Hobert, J. P., & Marchev, D. (2008). A theoretical comparison of the data augmentation, marginal augmentation and px-da algorithms. *Annals of Statistics*, *36* (2), 532–54.

Homburg, C., Koschate, N., & Hoyer, W. D. (2006). The role of cognition and affect in the formation of customer satisfaction: A dynamic perspective. *Journal of Marketing*, *70* (3), 21–31.

Imai, K., & Dyk, D. A. van. (2005). A bayesian analysis of the multinomial probit model using marginal data augmentation. *Journal of Econometrics*, *124* (2), 311–34.

Lenk, P., DeSarbo, W. S., Green, P. E., & Young, M. R. (1996). "hierarchical bayes conjoint analysis: Recovery of partworth heterogeneity from reduced experimental designs". *Marketing Science*, *15* (2), 173–91.

Lenk, P., & Orme, B. (2009). The value of informative priors in bayesian inference with sparse data. *Journal of Marketing Research*, *46* (6), 832–45.

Liechty, J. C., Fong, D. K. H., & DeSarbo, W. S. (2005). Dynamic models incorporating individual heterogeneity: Utility evolution in conjoint analysis. *Marketing Science*, *24* (2), 285–93.

Liu, X., & Daniels, M. (2006). A new efficient algorithm for sampling a correlation matrix based on parameter expansion and re-parameterization. *Journal of*

Computational and Graphical Statistics, *15*(4), 897–914.

Martin, A. D., & Quinn, K. M. (2002). Dynamic ideal point estimation via markov chain monte carlo for the u.s. supreme court, 1953-1999. *Political Analysis*, *10*, 134–53.

McCulloch, R., Polson, N. G., & Rossi, P. E. (2000). A bayesian analysis of the multinomial probit model with fully identified parameters. *Journal of Econometrics*, *99*, 173–93.

McCulloch, R., & Rossi, P. E. (1994). An exact likelihood analysis of the multinomial probit model. *Journal of Econometrics*, *64*, 207–40.

Meuwissen, T. H. E., & Goddard, M. E. (2004). Mapping multiple qtl using linkage disequilibrium and linkage analysis information and multitrait data. *Genet Sel Evol*, *36*(3), 261-79.

Mittal, V., Anderson, E. W., Sayrak, A., & Tadikamalla, P. (2005). Dual emphasis and the long-term financial impact of customer satisfaction. *Marketing Science*, *24*(4), 544–55.

Montgomery, A. L. (1997). Creating micro-marketing pricing strategies using supermarket scanner data. *Marketing Science*, *16*(4), 315–37.

Morgan, N. A., & Rego, L. L. (2006). The value of different customer satisfaction and loyalty metrics in predicting business performance. *Marketing Science*, *25*(5), 426–39.

Narasimhan, C., & Zhang, Z. J. (2000). Market entry strategy under firm heterogeneity and asymmetric payoffs. *Marketing Science*, *19*(4), 313–27.

Newton, M. A., & Raftery, A. E. (1994). Approximate bayesian inference with the weighted likelihood bootstrap. *Journal of the Royal Statistical Society Series B*, *56*(1), 3–48.

Nobile, A. (1998). A hybrid markov chain for the bayesian analysis of the multinomial probit model. *Statistics and Computing*, *8*(3), 229–42.

Oh, M.-S., & Raftery, A. E. (2001). Bayesian multidimensional scaling and choice of dimension. *Journal of the American Statistical Association*, *96*, 1031–44.

O'Hara, R. B. O., & Sillanp, M. J. (2009). A review of bayesian variable selection methods: What, how and which. *Bayesian Analysis*, *4*(1), 85–118.

Park, J., DeSarbo, W. S., & Liechty, J. (2008). A hierarchical bayesian multidimensional scaling methodology for accommodating both structural and preference heterogeneity. *Psychometrika*, *73*(3), 451-72.

Rossi, P. E., McCulloch, R. E., & Allenby, G. M. (1996). The value of purchase history data in target marketing. *Marketing Science*.

Rousseeuw, P., & Molenberghs, G. (1994). The shape of correlation matrices. *The American Statistician*, *48*, 276–79.

Shin, J., & Fong, K., Duncan K. H. nd Kim. (1998). Complexity reduction of a house of quality chart using correspondence analysis. *Quality Management Journal*, *5*, 46–58.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Linde, A. van der. (2002). Bayesian measures of model complexity and fit. *Journal of The Royal Statistical Society Series B-statistical Methodology*, *64*, 583–639.

Takane, Y., Young, F., & Leeuw, J. (1977). Nonmetric individual differences multidimensional scaling: An alternating least squares method with optimal scaling features. *Psychometrika*, *42*(1), 7-67.

Tonellato, S. F. (2001). A multivariate time series model for the analysis and prediction of carbon monoxide atmospheric concentrations. *Journal of The Royal Statistical Society Series C*, *50*(2), 187–200.

West, M., & Harrison, J. (1997). *Bayesian forecasting and dynamic models*. Springer-Verlag.

Yi, N., George, V., & Allison, D. B. (2003). Stochastic search variable selection for identifying multiple quantitative trait loci. *Genetics*, *164*(3), 1129–38.

Zhang, J., & Krishnamurthi, L. (2004). Customizing promotions in online stores. *Marketing Science*, *23*(4), 561–78.

Zhang, J., & Wedel, M. (2009). The effectiveness of customized promotions in online and offline stores. *Journal of Marketing Research*, *46*(2), 190–206.

Zhang, Y., & Koren, J. (2007). Efficient bayesian hierarchical user modeling for recommendation system. In *Sigir '07: Proceedings of the 30th annual international acm sigir conference on research and development in information retrieval* (pp. 47–54). ACM.

Vita
Zhe Chen

Education

Ph.D. in Statistics, The Pennsylvania State University, USA, 2014 (Expected)
B.S. in Mathematics, Fudan University, China, 2005
Exchange Student in Mathematics, The University of Hong Kong, Hong Kong, 2003

Professional and Teaching Experience

Student Instructor 05/2008 - 08/2008
Statistical Department in Penn State University, University Park, PA

Summer Intern 05/2007 - 08/2007
Penn State Milton S. Hershey Medical Center, Hershey, PA

Graduate Student Consultant 01/2007 - 12/2007
Statistical Consulting Center in Penn State University, University Park, PA

Publications

Duncan K. H. Fong, Sunghoon Kim, **Zhe Chen**, Wayne S. DeSarbo, (2014) “A Bayesian Multinomial Probit Model for the Analysis of Panel Choice Data”, *Psychometrika*

Zhe Chen, Duncan K. H. Fong and Wayne S. DeSarbo, “A Bayesian Vector Multidimensional Scaling with Model Selection Procedure for the Analysis of Rating Data” (working paper).

Rui Wang, Duncan K. H. Fong and **Zhe Chen** (equally contributed), “A Heterogeneous Bayesian Dynamic Model A for Replicated Cross-sectional Data” (working paper).

Conferences Presentations

Contributed Talk, *JSM,* Vancouver, Canada 08/2010
A New Hierarchical Bayesian Multinomial Probit Model for the Analysis of Consumer Choice Data

Contributed Poster, *Rao Prize Conference*, University Park, PA 06/2009
A Hierarchical Bayesian Dynamic Model for Customer Satisfaction and Shareholder Value Association Study